Quicker method for assessing influences on teaching assistant buy-in and practices in reformed courses

Matthew Wilcox, Yuehai Yang, and Jacquelyn J. Chini*

Department of Physics, University of Central Florida, Orlando, Florida 32816, USA

(Received 30 April 2016; published 5 August 2016)

Teaching assistants (TAs) that lead reformed recitations and labs must understand and buy into the design of the course and the research-based instructional strategies that the course requires in order to create high-fidelity implementations. We present a model that outlines possible influences on TAs' buy-in and their in-class actions coupled with a method, using a Real-time Instructor Observation Tool-based [E. A. West *et al.* Phys. Rev. ST Phys. Educ. Res. **9**, 010109 (2013)] exercise, to measure the effect of these influences that is not only quicker than interviews but also allows one to quantify these effects. We use this method to measure the influences on six graduate TAs teaching algebra-based introductory mechanics and electricity and magnetism recitations and labs ("mini studios") at the University of Central Florida. The results from the exercise are confirmed by interview responses from the TAs. We find a relatively high degree of buy-in to the design of the course, yet this is not reflected in the TAs' actions. The TAs' actions appear to be most influenced by student responses and expectations which do not align with the design of the course. Our study examines the effect of three influences shown in our model, and we argue that our method could be easily adapted to examine additional influences.

DOI: [10.1103/PhysRevPhysEducRes.12.020123](https://doi.org/10.1103/PhysRevPhysEducRes.12.020123)

I. INTRODUCTION

While active learning has been shown to increase student learning across the science, technology, engineering, and mathematics (STEM) fields [1], the course transformation process has still been slow [2]. Implementing active learning in lecture requires individual faculty to decide to transform their courses, and not all faculty are ready to do so. Instead, a department may try to ensure that all students have some opportunities to engage in active learning by transforming the recitation and lab section of an introductory physics sequence. Research has shown increased student performance on conceptual inventories after implementation of inquiry-based laboratories [3] and research-based tutorials [4]. At many large research universities, labs and recitations are led by graduate teaching assistants (TAs) [5]. This may be an advantage, as the department can require TAs to implement a uniform curriculum and students report finding TAs more relaxed, engaged, and relatable than faculty [6].

Just as instructor effectiveness has been linked to student performance, TAs who implement research-based instructional strategies (RBIS) can improve student performance on matched assessments compared to TAs who do not implement such strategies [7]. TAs with high-fidelity

implementation, defined as teaching with a high degree of similarity to the course designer's intentions, of the RBIS (for example, implementing suggested TA check-ins) show even greater improvement in student learning [8]. Since RBIS may be new to TAs, pedagogical training is especially important for TAs expected to lead transformed courses. This training may take the form of semester-long courses and/or weekly preparation meetings where TAs should discuss their roles, student questioning techniques, and facilitation skills [9]. For example, the designers of *Tutorials in Introductory Physics* required weekly preparation meetings where TAs would complete the worksheet and discuss common student mistakes [10]. Others have built on this model by pairing experienced and novice TAs to allow modeling of effective techniques [8]. Prior research has shown that an important job of TA preparation programs is to generate "buy-in" from the TAs for the reformed curriculum [11] as teachers' beliefs play a significant role in their instructional behavior [12]. These beliefs are complex, including general conceptions about teaching, students, and the makeup of a class [12].

How does one know if TA buy-in has been achieved? Prior research has used interviews to gauge belief-level buy-in and observations to determine practice-level buy-in [13–15]. In this study, we pair observations with the real-time instructor observation tool (RIOT) [16] with TA reports on a RIOT-based worksheet to gauge the effectiveness of a department's TA training system at generating TA buy-in. First, we demonstrate the utility of this worksheet-based method, which is simpler to implement than individual TA interviews, to both gauge TA buy-in and explore

*Corresponding author.
jchini@ucf.edu

Published by the American Physical Society under the terms of the *Creative Commons Attribution 3.0 License*. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

causes of failure to generate buy-in. Second, we find that while the department's combination of a first-semester TA pedagogy seminar and weekly preparation meetings effectively generates belief-level buy-in, TAs do not fully reach practice-level buy-in. Rather, our RIOT-based exercise reveals that student attitudes likely remain a major factor affecting TA practices.

II. LITERATURE REVIEW AND THEORETICAL FRAMEWORK

A. Graduate TAs need training to lead transformed courses

Prior studies have found that TAs do not feel prepared to lecture, facilitate discussions, or supervise students [17], and that many TAs do not feel their programs adequately prepare them to teach lecture, discussion, or laboratory courses [18]. Investigations of TAs' teaching practices within the same department implementing a common curriculum have revealed a wide range of pedagogical strategy uses [13,16,19]. For example, TAs enacting a student-centered curriculum demonstrated a wide variation in teaching actions, as coded by RIOT, despite TAs receiving "two intensive days of training, 10 weeks of continuing TA professional development, and extensive TA notes supporting a particular pedagogy" (Ref. [16], p. 10). Such variation may be expected when training programs do not take into account TAs' individual beliefs about teaching, as their chosen teaching practices likely align with their own understanding of how students learn [20]. For example, some TAs may not see the distinct roles a curriculum designer intends the lecture and recitation or discussion to fulfill [21], leading to frequent use of verification strategies within a curriculum that intends to emphasize student sense making [22]. When TAs' beliefs do not coordinate with the curriculum design, there is likely to be low alignment between their teaching practices and the curriculum design

[13–15]. Prior efforts have typically relied on individual interviews to assess the buy-in achieved by TA preparation programs, but these interviews are time-consuming for a department to conduct.

B. Many factors influence instructional practices

While instructors' beliefs may affect their teaching practices, this alignment is not complete and belief-level buy-in does not directly lead to practice-level buy-in [13–15]. Studies have found that instructors' discursive claims (statements about how their beliefs should be enacted in a classroom) more strongly correlate with their actual practices than do general statements about their beliefs [23]. This is likely because a variety of factors influence how instructors, including TAs, actually teach. We aggregate these influences to develop a theoretical model of influences on TA teaching practices, grouping influences into two main categories based on prior research (experience and systemic influences) and a third category more specific to the TA experience (class design). Figure 1 depicts these influences.

While the phrase "teachers teach how they were taught" is often repeated, this view overlooks the many other sources of professional knowledge and outside experiences that can influence teaching [24]. Sugrue finds that student teachers are also influenced by family, significant others, observations, atypical teaching episodes, and tacitly acquired understandings [25]. Personal views on teaching, prior everyday teaching experiences, and observations of teachers will affect how a TA views their role in the classroom. Oleson and Hora also find that teachers' own learning styles affect their instructional styles [24]. Thus, TAs enter their teaching assignment with prior experiences and beliefs that create a complex base [12] for TAs to build upon, and we must respect these experiential influences [20] if we hope to shape the TAs into someone who teaches using our intended RBIS.

FIG. 1. A schematic of the influences on a TA's teaching practices. The interaction of the experiential and class design influences is where initial buy-in is lost or gained. Initial buy-in may change based on systemic influences leading to a mediated buy-in.

Additionally, an instructor experiences systemic influences from the students, the department and university culture, as well as in-class limitations. Student attitudes and resistance are commonly cited as an influence on teaching and most frequently as a barrier to student-centered teaching [8,21,26–32]. TAs working in highly student-centered courses have expressed difficulty getting students to work in groups and generating student buy-in or maintaining motivation to participate in active learning during the full class period [32]. TAs decide how to shape the learning environment based on their observations of students [21]. Department and university norms may make it difficult to change away from traditional teacher-centered instruction [26,27] and for TAs to place significant value on teaching [33,34], possibly creating a “social and environmental context” [11] in which TAs are unlikely to buy into RBIS. In-class limitations include lack of time, the trade-off between depth and breadth in content coverage, and class size limitations [26–28]. Systemic influences can take precedence over the other influences. A TA may come into the classroom with complete buy-in, but systemic influences may prevent them from implementing the RBIS.

Unlike faculty, TAs assigned to labs and recitations may be expected to enact a predesigned curriculum, which may or may not expect them to use RBIS. While each TA has their core beliefs and experiences, how those are translated into teaching will be influenced by the design of their assigned course, including assumptions about what the TA and students should be doing during class time. This is an area of possible conflict with the TAs’ prior experiences as well as systemic influences. If the TA’s experiences and the course design are not aligned, the TA’s in-class actions may also be misaligned with the intended strategies. Thus, the course coordinator may become an advocate for the course design and an intermediary to help TAs interweave their experiences and the experience they are expected to create for students to generate buy-in. The centrality of the course coordinator within the department and university may influence how likely TAs are to be influenced by this design advocate. On the other hand, if a TA’s past experiences and the departmental and university culture align with the assigned course structure, we may expect more immediate buy-in. Using RIOT we cannot only quantify and determine the level of TA buy-in but also identify which aspects of the course a TA may not agree with. Our method also allows us to see what factors may be influencing a TA’s practices beyond their teaching beliefs.

III. METHODS

A. Context

We interviewed six TAs at the beginning and end of the Fall 2015 semester at the University of Central Florida (UCF), a very large, research intensive (Carnegie 2015 classification “highest research activity”) metropolitan

university. Of the six TAs, four were teaching a first-semester (mechanics) algebra-based mini studio while the other two TAs taught a second-semester [electricity and magnetism (E&M)] algebra-based mini studio. Three TAs were men and three were women, with all three men teaching the mechanics class. In both courses, half of the TAs interviewed were first-year graduate students at UCF and were taking a pedagogy seminar designed for physics graduate teaching assistants. The other half were in their second year of graduate school at UCF and took a similar pedagogy seminar during their first year. The graduate TA (GTA) pedagogy seminar was implemented after faculty found that their TAs were not as well trained in teaching methods as their undergraduate learning assistants. Thus, the GTA pedagogy seminar was modeled after the University of Colorado learning assistant pedagogy seminar [35], and engaged graduate students in reflecting on their teaching practices, evaluating research-based changes to courses, sharing their in-class experiences, and discussing education literature. The seminar explicitly focused on questioning, including Bloom’s taxonomy and the value of open-ended vs closed questions.

With support from a National Science Foundation grant, the physics department transformed the three-hour laboratory section of its algebra-based introductory physics sequence into a mini studio. The mini studio is composed of a 75-minute tutorial based on the University of Maryland Open Source Tutorials [36], followed by a 15-minute quiz and 80 minutes of lab based on the Investigative Science Learning Environment (ISLE) [37] curriculum. The mini-studio course structure has been shown to result in higher gains on the Force Concept Inventory compared to traditional and studio courses at UCF [19].

Every week of the semester, the mini-studio TAs attend a prep meeting for their specific course intended to prepare them to use the RBIS required to effectively lead the mini studios. Initially, these prep meetings exposed TAs to the ideas and tools of interactive and inquiry-based learning. In light of our model of the influences on a TA, this establishes a portion of the framework for the course. As TAs became more familiar with the RBIS, they were given opportunities to practice their implementation in the familiar prep meeting setting. The lab coordinator helped maintain the TAs’ understandings of the class design by allowing for such practice and giving the TAs an opportunity to reflect on their own teaching practice and their students’ engagement after using the RBIS. The prep meetings allow TAs to share and discuss their feelings about the class and the design with other TAs, which may spur them to change or expand their own teaching philosophies.

B. Observations

In addition to interviews, we observed the TAs throughout the semester and coded their actions using RIOT [16]. RIOT codes an instructor’s actions in real time using an

TABLE I. The details of each action in RIOT and what the TA was doing in the classroom when they were coded.

	Action	Description
Talking at students	Clarifying	Clarifying instruction, reading from worksheet
	Explaining	Explaining a concept without student input
Shared student-TA dialogue	Student question	Listening to a question
	Closed dialogue	TA controls conversation, students give short answer
	Open dialogue	No person is controlling the conversation, student responds with sentences
Observation	Passive observing	Listening briefly to group conversations
	Active observing	Listening to one group but does not participate
	Student presentation	Listening to students presenting work to class
	Students talking serially	Listening to students responding to each other in a whole class discussion
Not interacting	Not interacting	Reading notes, grading, preparing to teach

exhaustive and exclusive list of codes. Any action that an instructor may take corresponds to a single RIOT code. As such, at any point in time during instruction there is only one action coded. The RIOT protocol allows an observer to determine the percentage of time an instructor spends on each of ten actions that fall under four main categories, as described in Table I. We observed each TA three times during the semester, as shown in Fig. 2. The RIOT observations were performed by two researchers. Before collecting data, the two researchers discussed the RIOT codes to come to an agreement on what actions constitute which codes. One researcher was experienced in RIOT and taught the other how he coded and what he looked for. They then performed practice observations at the same time while coding individually. These were followed by a discussion of the observers' disagreements. This process was repeated until sufficient interrater reliability (IRR) was achieved as measured by a Cohen's κ greater than 80% with a tolerance of 6 seconds, as described in Appendix A.

C. Interviews and RIOT worksheet

During the postsemester interviews, we asked the TAs to fill out a worksheet (see Appendix B) in which they indicated the amount of time a TA would spend on the various teaching actions coded for in RIOT based on different perspectives: what the curriculum designer wants,

their own view of a helpful TA, what their students want, and a prediction of their actual teaching. In each case, we are asking the TA to interpret what they or an outside group think the TA should do during class time. The prompts for each perspective are listed in Table II. Once the TAs became familiar with the codes and their definitions we would prompt them to rank the amount of time spent on each action for a TA in different perspectives. The TAs created profiles by ranking the time spent on each action as either "none," "relatively low," "medium," or "relatively high." When the TAs had completed the worksheet, they were shown their actual teaching profiles and asked to say aloud anything that struck them as interesting or surprising. Additionally, TAs talked about their teaching plans in the presemester interview and their teaching experience in the postsemester interview.

IV. DATA ANALYSIS

We want to know how the TAs' responses change based on the different TA perspectives. To do so, we have to quantify the total number of differences in their rankings for each action in every possible comparison. We calculated a number, called δ , which is the number of times a TA disagreed in their responses to two different questions. A higher δ indicates larger disagreement in the profiles of two perspectives and a lower δ indicates a larger degree of

FIG. 2. The research timeline for interviews and observations.

TABLE II. The prompts for the TA profile perspectives that the TAs needed to fill in as given in the interviews.

Perspective	Prompt
Curriculum design	What would you estimate is the average amount of time a TA should spend on each of these actions, based on the design of the mini studio?
Helpful TA	What profile do you feel would be most helpful for the students?
Students	What do you think the students would like their TA's profile to be?
Predicted	What would you estimate is the average amount of time spent on each action while you were teaching this semester?

agreement. As an example, imagine a TA were to respond to the Design perspective with “medium” for all ten actions. Then, in the Helpful TA perspective, he responds “medium” to all actions, except he puts “relatively low” for Explaining and “relatively high” for Open Dialogue. For the Design vs Helpful TA comparison, this TA's δ would be 2 since there were only two actions where his response disagreed between these two questions. This treatment is done for all possible comparisons and for all TAs. For each comparison, we add the six TAs' δ s to get a total difference called Δ .

We asked the TAs to give their responses in this ordinal data type to make the worksheet less demanding. Because of the nature of the ordinal data type, it is incorrect to report averages and standard deviations of the responses as these are only appropriate for data types in which the distances between successive responses are equal, such as the interval or ratio data types. We cannot average across TAs because we cannot reasonably assume that the difference between, for example, “relatively low” and “medium” for one TA is the same as that for another TA. By counting the number of disagreements within a TA we avoid any issues with not having interval type data. If we were to assume that for an individual TA the difference between “relatively low” and “relatively high” means twice the difference between “medium” and “relatively high,” then we may be able to better gauge the magnitude of the disagreement between two perspectives. Although it may not be safe to make this assumption, it may be a more intuitive way to view the data. As such, we include this analysis in Appendix C. All results for this analysis agree with the results from our original analysis.

There were six different profiles consisting of the TAs' responses to the four prompts, their actual profile, and a profile of an ideal TA as agreed upon by the authors. We are familiar with the research-based approaches desired in the mini studios and have been lab coordinators for and/or have done research in this style of class. We separately created a profile for an ideal TA based on the design of the course then came together, discussed, and agreed upon a final

profile. This profile is shown in Table III. Across these six profiles we found the Δ for all 15 pairwise comparisons.

A. Analyzing observation data

To compare a TA's responses to one question to their actual profile, we must first transform their actual profile into the rankings of “none,” “relatively low,” etc. Figure 3 displays the average percentage of time one of our six TAs spent on each action across all three observations. The shaded area is the portion of the graph that is equal to the average of all the data plus or minus half of the standard deviation of averages for each action. The thickness of the shaded area is unique to each TA due to different standard deviations, but it is always centered on 10% (since the average percentage of time spent on ten actions is 10%). Actions with a percentage of time spent that falls within the shaded area are categorized as having spent a “medium” amount of time doing that action. Actions that fall above or below this area are coded as “relatively high” or “relatively low.” If an action did not occur in any observation, that action is categorized as “none.”

B. Within-group comparisons

Looking at the group of comparisons that contain one profile allows us to make a claim about that one profile. For example, if we want to find which profile is most similar to what the TAs think a TA should do based on the design of the course, we find the Δ for the Design vs Helpful TA, Design vs Student, Design vs Prediction, Design vs Actual, and Design vs Agreed Design comparisons. We find that the smallest Δ in this group corresponds to the Design vs Helpful TA profile, and so we can claim that the TAs' conception of how a TA should act based on the design of the course is most similar to what the TAs think is most helpful in the course. Generally, these comparisons are not reciprocal. Suppose the most similar comparison in the Student group is Student vs Predict. That does not necessarily mean that the most similar comparison in the Predict group is Predict vs Student.

TABLE III. The profile of an ideal TA based on the design of the curriculum as agreed upon by the authors.

Clarifying	Explaining	Listening to question	Closed dialogue	Open dialogue	Passive observing	Active observing	Student presentation	Students talking serially	Not interacting
Low	Low	Medium	Low	High	Medium	Medium	Medium	Medium	None

FIG. 3. The RIOT profile for a single TA. Data which fall below the shaded area are categorized as “relatively low.” Data above the shaded area are categorized as “relatively high.” Data within the shaded area are categorized as “medium.” For this TA, there was no occurrence of the Students Talking Serially code, so it would be categorized as “none.”

V. RESULTS

The average percentage of time spent on each action coded in RIOT across all TAs is shown in Fig. 4. From the graph, we see that the average TA spent nearly half of their class explaining or not interacting and nearly 4% of their class engaging in open dialogue. This is similar to prior findings with RIOT where TAs have been found to spend 45% of their time in a lab or workshop not interacting with the students [38]. Based on the design of the class we expect less time spent explaining, more time spent in open dialogue, and no time at all not interacting. To make claims about the reasons for the high prevalence of explaining and not interacting, we make comparisons between the TAs’ responses to each perspective prompt and their actual RIOT profiles. The results disaggregated by gender are presented in the Supplemental Material [39] and reveal no major difference in how they taught their classes.

A. Comparisons

Figure 5 shows the most similar of the within-group comparisons. We excluded the Student and Predict groups for brevity since we draw no conclusions from those within-group comparisons. Figure 5 also indicates how many TAs agree with the consensus for the most similar within-group comparison. If a TA’s portion of a bar is striped, that indicates that the comparison was not the most similar comparison for that specific TA. For many of the most similar comparisons, at least half of the TAs agree with the consensus. For all but one comparison, the most similar comparisons reported have a Δ that is at least 20% smaller than the second most similar comparison in that group, indicating that the most similar comparison is substantially better than the next in that group. The only exception to this is in the Actual comparison group, which is discussed next.

FIG. 4. Average of all TAs’ actual profiles as coded by RIOT.

FIG. 5. A representation of the Δ 's for each group of comparisons. Only the smallest Δ 's for select groups are shown. Two Δ 's are shown for the "Actual" group due to the two comparisons having similar Δ 's (39 and 40). Each section separated by color in each bar represents a TA's δ . If that section is striped it indicates that this was not the most similar comparison for that group for that TA.

1. TAs demonstrate buy-in to course design

Design vs Helpful TA was the most similar comparison for both the Design group and the Helpful TA group. In fact, this was the most similar comparison across all possible comparisons with a Δ of 24. This indicates that the TAs are buying into the design of the course. Not only are the TAs buying into what they consider to be the design of the course, but the TAs are also understanding the design of the course to a relatively high degree of similarity. With a Δ of 30, the TAs' ideas of how a TA should act based on the design of the course and our ideas are similar. The next most similar profile in the Agreed Design group is Helpful TA with a Δ of 38. This shows that it is likely that the course coordinator is effectively conveying the ideas behind the design of the class.

TAs' discussions in the interviews support the conclusion that they bought into the course design. All six TAs described their role in the mini studio as that of *facilitator*. When asked how he planned to teach the mini studio that semester, TA-C stated, "I think my role is always to facilitate the students, to give hints and to actively encourage them in learning." Similarly, when asked to describe her role as a TA in the mini studio, TA-E explained, "so what I'm trying to do is being there for the students to somehow guide them through the way they're going and to give them some clues about what's going on in the class. But, like, they ask a lot of questions during the lab, I try to somehow indirectly tell them which way to go but just not giving the exact solution and answers so they can try to figure out what's going on the class. It's more like a leader or guider than somebody who tells them what to do." Additionally, several TAs described ways they changed their teaching strategies to better align with the mini-studio course design. When describing how he ran his mini-studio

sections, TA-A described implementing changes based on both the weekly preparation meeting and the pedagogy seminar: "I tried to sort of change it up every now and again to see what worked best, especially what we were talking about in the prep meetings. So, a few of them I would do the majority of the writing on the board and I would sort of get input from the groups to help me out there. Or, in the middle of the semester I passed out markers and had them go through the worksheet in little chunks amongst themselves and write notes on the board... I think I started observing more over the course of the, over the semester. There was a lot of talk in the seminar about giving them time to think and just making your presence known in general..." Similarly, TA-D described trying new strategies in her second semester of leading a mini studio: "So I would say in regards to the spring semester last year it was a new concept to me, mini studios. So I guess I was trying to apply more of the techniques that I practiced before, which was more like lecture mode, and this time around I actually gave up a lot of those techniques. I let the students actually take control of what they were doing more so."

2. Buy-in does not directly lead to high-fidelity implementation

Does this buy-in translate into practice? Our observations prove that this is not necessarily the case as the TAs' actions do not demonstrate as much buy-in as do their beliefs. The Δ 's for Actual vs Design, Actual vs Agreed Design, and Actual vs Helpful TA are 48, 47, and 45, respectively. These are the three largest Δ 's out of all possible comparisons. Although the TAs are buying into how we want them to teach the class, there appears to be another larger influence on their practices. Aside from their own predictions, what the TAs actually did in their classes

most resembles what they feel the students want from their TAs. This comparison has a Δ of 40. With a Δ of 42, what the TAs think the students want from their TAs is least similar to how the TAs believe the class should be taught based on the design. We can see that although the TAs buy into the design of the class, their actions are heavily influenced by what they think the students want and what the students want is least similar to the TAs' ideas of the design of the course.

In the interviews, several TAs described students requesting changes to their initial teaching style and in some cases making the requested changes. For example, TA-E stated, "The way I taught at first in the class is completely different than the way it came out with the students. What makes me be less interactive with the students, maybe some of the things, like active observing, they don't like it. I'm not saying I shouldn't do it because they don't like it but this is the feedback I get from them... Sometimes they give me the feedback like they were more confused when I asked questions about what they were asking or something like making an open dialogue so they can discuss about the things. They just want a straight forward answer and when you don't give them that they won't ask you so your chance of talking to them is less..." This TA expressed the concern that sticking to her intended use of open dialogue may actually shut down her opportunities to have any type of dialogue with the students.

VI. DISCUSSION

A. Sources of TA buy-in

Using our RIOT-based exercise as an analysis tool for our TAs, we are able to determine that they know how we want them to teach and agree that this is useful (the TAs buy into the reformed course); nearly all TAs had Agreed Design vs Design as their most similar comparison in the Agreed Design group, further proving that the TAs understand the course design. Our current technique does not reveal how the TAs developed this buy-in, but several quotes from our interviews suggest that the weekly preparation meeting and pedagogy seminar may have worked in tandem to convey a sense of a departmental expectation of student-centered teaching, contributing to a supportive "social and environmental context" [11]. For example, TA-C described coming to realize the expectations in this department were different from those he had experienced in his undergraduate education and his prior teaching experiences outside the U.S. "The first semesters I came here I tended to do as I did in [country of origin]. Like lecturing as much as I can. And I gradually change my teaching style as the department force students and TAs to actively involve students in active learning." Later in the interview, this TA indicates some of the influences that created this view: "... we had, like, this pedagogy class the first semester we came here. They wants, and I think the whole department

wants TA to facilitate students solving the problems. Not giving them the exact answer. So the only thing they want, I think, is giving some help if they need the help." Referring back to our theoretical model (Fig. 1), TA-C had an experiential influence from his prior teaching. Through the pedagogy seminar, this TA also acquired a systemic influence from the departmental culture and expectations. It appears that the systemic influence eventually won over the experiential influence and has become a major factor in the TA's teaching style. TA-A similarly described the course transformation as a departmental undertaking: "I would assume the department is looking to improve on the lab structure." Thus, the TAs appear to have a sense that active learning is not just expected by the course coordinator, but rather by the larger department, and it seems that sense may have been fostered by departmental required training, such as the weekly preparation meeting and the first year pedagogy seminar. Prior comparison between departments with varying degrees of TA buy-in for research-based tutorial instruction has suggested that TA buy-in is likely enhanced by the perception that tutorial instruction is "part of the accepted department practice" [11].

Additionally, some of our TAs have had a wide variety of learning experiences due to the national increase in active learning and may use their positive experiences with student-centered instruction as the model of teaching they would like to emulate. When asked to describe the type of courses he had taken as an undergraduate, TA-A described that one of his upper-level physics courses used some "research-based techniques." He described, the instructor "had his little lecture, and then we'd break off into groups... we would work on these worksheets together and he would come around to see what we were up to and give us some guidance if he felt it was necessary... There was lecture and there was a professor teaching and an LA but having worksheets and having a little bit more time to really think about them critically in the lecture and use that time for something other than copying down notes from the board, that was pretty beneficial." The theoretical model for TA-A would include prior active-learning experience as an experiential influence. His favorable view of this experience leads to buy-in. His understanding of the departmental values, a systemic influence, aligning with active learning supports his buy-in. Thus, one way to generate TA buy-in for active learning in introductory courses may be to provide positive experiences with active learning in upper-level physics courses and a department that supports those values.

B. Training should focus on generating student buy-in

Despite exhibiting strong belief-level buy-in, the TAs in this study still do not fully put their student-centered views into practice. Prior research has suggested many "barriers" to student-centered instruction [8,21,26–32]. Our analysis puts these barriers in a new light by indicating that TAs

seem to align their teaching strategies with the strategies they think their students expect, in some cases even to a greater degree than the TAs expect, as evidenced by TA's Actual RIOT profiles (based on in-class observations) most closely matching their Predicted profiles and Student profiles.

As demonstrated, the TAs in this study described instances of students demonstrating resistance through such actions as demonstrating dissatisfaction, suggesting alternate teaching methods, and resisting participation in TAs' attempts to implement active learning techniques, which align with some of the examples of student resistance compiled by Seidel and Tanner [40]. For some TAs, the training provided in the pedagogy seminar and the weekly training meetings provided a sort of "buffer" against students' displeasure. TA-F described that the pedagogy seminar affected her teaching by making her "just take that step back and letting them talk instead of me. It's very difficult sometimes when they get frustrated with you for not giving them the answer, but I think the seminar helped me see the value in that." TA-D similarly shared, "I've had students actually tell me a couple of times to do a lecture thing and I've had to tell them many times that I cannot do it and it's against what mini studio actually stands for. But, I've had to explain to them, too, because at the end of the day I was more concerned about them having the correct concept about what the subject is about. We had a good discussion about this in our [preparation meeting]... [lab coordinator] even mentioned, it's not really important for us to give them the right concept... it's more important for us to let them think, and I did have a good discussion with him about that." Thus, both training experiences appear to buffer TAs' teaching beliefs against students' expectations, although our observations suggest their actions are still influenced by the TAs' perceptions of their students' concerns. In our theoretical model, the situation

described by TA-D would appear as the systemic influence of student expectations and response blocking the buy-in. TA-D may exhibit appropriate buy-in, but ultimately the students' responses remained a larger influence on her practices. Prior research has found that TAs who view their students as "knowledge builders" were more likely to report using active learning [30], and that TAs highly value discussion of student expectations during preparation meetings [17]. With a Δ of 42 for both, the difference between Student vs Design and Student vs Agreed Design is the largest difference in the Student group, indicating that our TAs perceive their students to expect different teaching strategies than those expected by the course coordinator. However, the goals of the mini studio are well aligned with students' likely expectations, which may include individual and group problem solving, investigations of how things work, and classwide discussions [41]. Our training may need to focus more specifically on helping TAs communicate how RBIS can better achieve these goals compared to traditional instruction.

C. RIOT-based worksheet allows simple measurement of TA buy-in and barriers

Revisiting our theoretical model (see Fig. 6), we find that our RIOT-based worksheet allows us to capture many of the influences on TAs' teaching practices as proposed by prior research. The Design prompt probes how well a TA understands the design of a class, revealing the extent to which the course coordinator has shaped the TAs' views of the expected teaching practices. The Helpful TA prompt probes the experiential influences on a TA, revealing what they personally feel is the most helpful way to spend their time in the mini studio. Comparing the responses to the Design prompt and the Helpful TA prompt gives us a clear indication of the amount of buy-in there is for a TA

FIG. 6. The theoretical model with the aspects that we have measured. Aspects in gray could be measured given the appropriate prompt. Aspects that could be measured are not limited to those shown here. The legend indicates which prompt or comparison measures the corresponding influence.

(assuming the Design prompt responses are well aligned with the Agreed Design, as they are in this study). The Student prompt highlights the effect of this particular systemic influence. In the present study, we find that this may be the largest influence on our TAs' actions, despite their high level of buy-in.

D. Future work

Additionally, we would like to explore other perspectives through which TAs could be asked to consider their teaching strategies to further illuminate the conflicting vs supportive expectations within which they construct their teaching practices. New prompts could be used to focus on other systemic influences. For example, to explore how time limitations affect a TA, one could ask, "What would be the profile of a TA in this class who had as much time as they needed?" To probe the departmental influence, one could ask, "What do you think your department would like the profile of their TAs to be?" To probe particular prior experiences, one could ask, "What was the profile of your favorite undergraduate science course instructors?" or "Based on your observations of an expert teacher, what do you think is the best profile for an instructor in this course?" With carefully constructed prompts, we hypothesize that one could measure the degree of most proposed influences.

We have proposed simple models of the generation of TA buy-in. However, the present study explores a single curriculum in a single department. These models should be enriched and substantiated by exploring TAs' experiences in a wide variety of contexts.

VII. CONCLUSIONS AND IMPLICATIONS

We have developed a quick method of measuring TA buy-in and performance without the need for structured interviews. Using RIOT and having TAs predict RIOT profiles on a worksheet based on varying perspectives is an easy way to find out if TAs agree with the design of a course and if that buy-in translates into practice. Variations on this method can also reveal other influences on the TA's practices. These investigations generate data that one can use to tailor weekly prep meetings to address the TAs needs, which may include TA buy-in, practice with the instructional strategies, or helping TAs gain student buy-in. This method could also be easily expanded to audiences outside of physics TAs. The model and method were not specifically altered to accommodate a physics background or graduate students. As such, we believe this method may be used with undergraduate teaching assistants and instructors in other disciplines. In fact, we have used this method with mathematics GTAs at UCF with no difficulty.

In the context of mini-studio instruction at a department with a first-year TA pedagogy course and weekly course-specific preparation meetings, we find that our TAs do buy into the design of the course. This is evident by similar TA

responses to RIOT profiles from the perspective of the course designer (Design) and their own perspective (Helpful TA). Further, the TAs' understanding of the design of the class agrees with the profile of an ideal TA as agreed upon by us. However, despite sufficient buy-in, the TAs' actual profiles are least similar to the design of the class from both the TAs' perspective and our perspective. Other research has shown that TA buy-in can exist and not lead to aligned practices [11,14]. This may be due to barriers to the implementation of RBIS, such as student resistance and time management [28,31,32]. In our data, the most salient barrier is that of the TAs' conceptions of student expectations. Second only to their predictions, what the TAs actually did in their class (Actual) is most similar to what they felt students want from their TAs (Student). These results are substantiated by TAs' discussion in postsemester interviews, and together suggest that a supportive departmental culture may lead to TA buy-in, but is not enough to ensure high-fidelity implementation of RBIS in the presence of resistant student attitudes. The results of this research have led us to begin looking at training instructors to generate student buy-in so as to eliminate the pressure to change their teaching methods early in the semester.

ACKNOWLEDGMENTS

The authors thank physics graduate teaching assistants for their participation. This work was supported in part by National Science Foundation Grant No. DUE1246024.

APPENDIX A: CALCULATING COHEN'S κ

To calculate Cohen's κ , we determined what each observer was coding every 10 seconds throughout the observation. The value of 10 seconds was chosen since that is the minimum time required for an action to be coded in RIOT. If we chose a value larger than 10 seconds, then we may overlook a coded action in our IRR calculation; choosing lower values inflates the number of disagreements used in calculating Cohen's κ . We compared what one observer coded at one point in time with what the other observer had coded at that same point in time plus or minus the 10-second interval so that we do not count a disagreement when the two coders take different amounts of time to press a code. This can occur when the coders have slightly different cues for when a new action has started and when they count to 10 seconds at different rates. These differences, although minor, can have a significant impact on Cohen's κ . Using these variables to calculate Cohen's κ , we can then alter the time interval to determine the level of tolerance at which the coders agree. Our two observers agreed to a Cohen's κ of 80% with a 6-second tolerance.

APPENDIX B: RIOT-BASED SURVEY

Figures 7 and 8 are copies of the worksheet and RIOT code definitions that were given to TAs. The

Please review the separate definition sheet. Feel free to ask if you would like clarification of any of these definitions.

The questions below will ask you to describe how you do or could spend your time during recitation (across the categories described above) from several perspectives. For each question, write the **relative amount of time** that a TA in that perspective would spend on each action. You may respond with the following choices: an amount of time that is “Relatively High”, “Relatively Medium”, “Relatively Low”, or “No Time”.

Q1: How do you think the **course instructor** would like GTAs to spend their time during the recitation?

Clarifying Instructions	Explaining Content	Student Question	Closed Dialogue	Open Dialogue	Passive Observing	Active Observing	Student Presentation	Students Talking Serially	Not Interacting

Q2: How do **you** think is the **most helpful** way for a GTA to spend their time during the recitation?

Clarifying Instructions	Explaining Content	Student Question	Closed Dialogue	Open Dialogue	Passive Observing	Active Observing	Student Presentation	Students Talking Serially	Not Interacting

Q3: How do you think the **students in this course** want a GTA to spend their time during the recitation?

Clarifying Instructions	Explaining Content	Student Question	Closed Dialogue	Open Dialogue	Passive Observing	Active Observing	Student Presentation	Students Talking Serially	Not Interacting

Q4: How do you think **you actually** spend your time during the recitation?

Clarifying Instructions	Explaining Content	Student Question	Closed Dialogue	Open Dialogue	Passive Observing	Active Observing	Student Presentation	Students Talking Serially	Not Interacting

FIG. 7. A copy of the RIOT-based worksheet given to the TAs.

worksheet could be modified to have TAs respond with percentages of time in place of the ordinal time intervals used in this study.

Actions Coded in RIOT

	Action	Description
Talking At Students	Clarifying	Clarifying instruction, reading from worksheet
	Explaining	Explaining a concept without student input
Shared Student/TA Dialogue	Student Question	Listening to a question
	Closed Dialogue	TA controls conversation, students give short answer
	Open Dialogue	No person is controlling the conversation, student responds with sentences
Observation	Passive Observing	Listening briefly to group conversations
	Active Observing	Listening to one group but does not participate
	Student Presentation	Listening to students presenting work to class
	Students Talking Serially	Listening to students responding to each other in a whole class discussion
Not Interacting	Not Interacting	Reading notes, grading, preparing to teach

FIG. 8. A copy of the definitions for RIOT codes given to the TAs prior to completing the worksheet.

APPENDIX C: ASSUMING EQUAL DIFFERENCE ANALYSIS

By coding each response as a number, 0 for “none” up to 3 for “relatively high,” we can compute the difference in a response for one action between two perspectives. For example, in the question based on the design of a class a TA may have put “relatively low” for the Explaining action. This is coded as a 1. In the question about what they think the students find most helpful, that same TA may have put “relatively high” for Explaining. This is coded as a 3. Thus, the difference in Explaining in the design and student comparison is 2. Adding up these values across all 10 actions gives a number analogous to δ that will be called δ^* . Adding the δ^* s for the same comparison across all TAs gives us a number analogous to Δ called Δ^* .

Since the intervals between the response rankings may mean different things to different TAs, a δ^* of 10 for one TA could be equivalent to a δ^* of 8 for another TA. However, this effect is washed out when we compare Δ^* s. Imagine each δ^* as different world currencies with certain exchange rates between each TA. In the earlier example the exchange rate

FIG. 9. Most similar comparisons for the Design, Helpful TA, Actual, and Agreed Design groups. A striped section means that this is not the most similar comparison for that TA.

between the two TAs would be 10 to 8. Within one Δ^* , a TA may only contribute 80% of what another TA contributes, but that “exchange rate” is true across all Δ^* s. Thus, we can safely compare multiple Δ^* s.

We do make the assumption that within each TA the difference between rankings is equivalent. That is to say that the difference between “relatively low” and “relatively medium” is the same as the difference between “relatively

medium” and “relatively high” for one TA. If this were untrue, then we cannot say that the difference between “relatively low” and “relatively high” is twice the difference between “relatively medium” and “relatively high.” Nor can we say that the difference between “none” and “relatively medium” is the same as the difference between “relatively low” and “relatively high.” Accepting this assumption, we still get similar results, as shown in Fig. 9.

- [1] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, Active learning increases student performance in science, engineering, and mathematics, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8410 (2014).
- [2] National Research Council Committee on Undergraduate Physics Education Research and Implementation, *Adapting to a Changing World: Challenges and Opportunities in Undergraduate Physics Education* (National Academies Press, Washington, DC, 2013).
- [3] B. Thacker, H. Dulli, D. Pattillo, and K. West, Lessons from a large-scale assessment: Results from conceptual inventories, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020104 (2014).
- [4] S. J. Pollock and N. D. Finkelstein, Sustaining educational reforms in introductory physics, *Phys. Rev. ST Phys. Educ. Res.* **4**, 010110 (2008).
- [5] E. Seymour, *Partners in Innovation: Teaching Assistants in College Science Courses* (Rowman and Littlefield, Boulder, 2005).
- [6] K. D. Kendall and E. E. Schussler, Does instructor type matter? Undergraduate student perception of graduate teaching assistants and professors, *CBE Life Sci. Educ.* **11**, 187 (2012).
- [7] L. Deslauriers, E. Schelew, and C. Wieman, Improved learning in a large-enrollment physics class, *Science* **332**, 862 (2011).
- [8] K. M. Koenig, R. J. Endorf, and G. A. Braun, Effectiveness of different tutorial recitation teaching methods and its implications for TA training, *Phys. Rev. ST Phys. Educ. Res.* **3**, 010104 (2007).
- [9] J. A. Gilreath and T. F. Slater, Training graduate teaching assistants to be better undergraduate physics educators, *Phys. Educ.* **29**, 200 (1994).
- [10] L. C. McDermott and P. S. Shaffer, *Tutorials in Introductory Physics* (Prentice-Hall, Englewood Cliffs, NJ, 1998).
- [11] R. M. Goertzen, R. E. Scherr, and A. Elby, Accounting for tutorial teaching assistants’ buy-in to reform instruction, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020109 (2009).
- [12] A. G. Thompson, The relationship of teachers’ conceptions of mathematics and mathematics teaching to instructional practice, *Educ. Stud. Math.* **15**, 105 (1984).
- [13] R. M. Goertzen, R. E. Scherr, and A. Elby, Tutorial teaching assistants in the classroom: Similar teaching behaviors are supported by varied beliefs about teaching and learning, *Phys. Rev. ST Phys. Educ. Res.* **6**, 010105 (2010).
- [14] B. T. Spike and N. D. Finkelstein, Design and application of a framework for examining the beliefs and practices of physics teaching assistants, *Phys. Rev. Phys. Educ. Res.* **12**, 010114 (2016).

- [15] B. T. Spike and N. D. Finkelstein, Examining the beliefs and practice of teaching assistants: Two case studies, *AIP Conf. Proc.* **1289**, 309 (2010).
- [16] E. A. West, C. A. Paul, D. Webb, and W. H. Potter, Variation of instructor-student interactions in an introductory interactive physics course, *Phys. Rev. ST Phys. Educ. Res.* **9**, 010109 (2013).
- [17] G. Marbach-Ad, K. L. Schaefer, B. C. Kumi, L. A. Friedman, K. V. Thompson, and M. P. Doyle, Development and evaluation of a prep course for chemistry graduate teaching assistants at a research university, *J. Chem. Educ.* **89**, 865 (2012).
- [18] C. M. Golde and T. M. Dore, *At Cross Purposes: What the Experiences of Today's Doctoral Students Reveal About Doctoral Education* (The Pew Charitable Trusts, Philadelphia, 2001).
- [19] J. Chini and J. Pond, Comparing traditional and studio courses through FCI gains and losses, in *Proceedings of the Physics Education Research Conference, Minneapolis, 2014* (2016), pp. 51–54.
- [20] R. M. Goertzen, R. E. Scherr, and A. Elby, Respecting tutorial instructors' beliefs and experiences: A case study of a physics teaching assistant, *Phys. Rev. ST Phys. Educ. Res.* **6**, 020125 (2010).
- [21] J. A. Luft, J. P. Kurdziel, G. H. Roehrig, and J. Turner, Growing a garden without water: Graduate teaching assistants in introductory science laboratories at a doctoral/research university, *J. Res. Sci. Teach.* **41**, 211 (2004).
- [22] R. M. Goertzen, R. E. Scherr, A. Elby, C. Henderson, M. Sabella, and L. Hsu, Indicators of understanding: What TAs listen for in student response, *AIP Conf. Proc.* **1064**, 119 (2008).
- [23] W. R. Veal, M. E. Riley Lloyd, M. R. Howell, and J. Peters, Normative beliefs, discursive claims, and implementation of reform-based science standards, *J. Res. Sci. Teach.*, doi:10.1002/tea.21265 (2015).
- [24] A. Oleson and M. T. Hora, Teaching the way they were taught? Revisiting the sources of teaching knowledge and the role of prior experience in shaping faculty teaching practices, *Higher Educ.* **68**, 29 (2014).
- [25] C. Sugrue, Student teachers' lay theories and teaching identities: Their implications for professional development, *Eur. J. Teach. Educ.* **20**, 213 (1997).
- [26] M. H. Dancy and C. Henderson, Beyond the individual instructor: Systemic constraints in the implementation of research-informed practices, *AIP Conf. Proc.* **790**, 113 (2005).
- [27] C. Henderson and M. H. Dancy, Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics, *Phys. Rev. ST Phys. Educ. Res.* **3**, 020102 (2007).
- [28] C. Turpen, M. Dancy, and C. Henderson, Perceived affordances and constraints regarding instructors' use of peer instruction: Implications for promoting instructional change, *Phys. Rev. Phys. Educ. Res.* **12**, 010116 (2016).
- [29] N. U. Bautista, E. E. Schussler, and S. M. Rybczynski, Instructional experiences of graduate assistants implementing explicit and reflective introductory biology laboratories, *Int. J. Sci. Educ.* **36**, 1184 (2014).
- [30] T. A. Pinder-Grover, Active learning in engineering: Perspectives from graduate student instructors, in *Proceedings of the 120th Annual Conference of the American Society for Engineering Education, Atlanta, GA* (2013), http://www.asee.org/file_server/papers/attachment/file/0003/3855/TA_Training_Active_Learning_ENGR_EDU_AUDIENCE_3April13_n3.pdf.
- [31] E. Yerushalmi, C. Henderson, K. Heller, P. Heller, and V. Kuo, Physics faculty beliefs and values about the teaching and learning of problem solving. I. Mapping the common core, *Phys. Rev. ST Phys. Educ. Res.* **3**, 020109 (2007).
- [32] G. DeBeck, S. Settelmeyer, S. Li, and D. Demaree, TA beliefs in a SCALE-UP style classroom, *AIP Conf. Proc.* **1289**, 121 (2010).
- [33] A. E. Austin, Preparing the next generation of faculty: Graduate school as socialization to the academic career, *J. Higher Educ.* **73**, 94 (2002).
- [34] J. D. Nyquist, The PhD a tapestry of change for the 21st Century, *Change* **34**, 12 (2002).
- [35] V. Otero, S. Pollock, and N. Finkelstein, A physics department's role in preparing physics teachers: The Colorado learning assistant model, *Am. J. Phys.* **78**, 1218 (2010).
- [36] A. Elby *et al.*, Open Source Tutorials in Physics Sense-making: Suite II (2008), <http://www.spu.edu/depts/physics/tcp/tadevelopment.asp>.
- [37] E. Etkina and A. V. Heuvelen, Investigative science learning environment: Using the processes of science and cognitive strategies to learn physics, in *Proceedings of the 2001 Physics Education Research Conference*, edited by S. Franklin, J. Marx, and K. Cummings (PERC, Rochester, NY, 2001), pp. 17–20.
- [38] E. M. Hickok, Master's thesis, San Jose State University, 2016, arXiv:1602.07740.
- [39] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.12.020123> for an analysis of the RIOT profile and survey responses by gender.
- [40] S. B. Seidel and K. D. Tanner, "What if students revolt?"—Considering student resistance: Origins, Options, and Opportunities for investigation, *CBE Life Sci. Educ.* **12**, 586 (2013).
- [41] J. D. Gaffney, Education majors' expectations and reported experiences with inquiry-based physics: Implications for student affect, *Phys. Rev. ST Phys. Educ. Res.* **9**, 010112 (2013).