

Impact of changing physical learning space on GTA and student behaviors

Constance M. Doty¹

¹*Department of Physics, University of Central Florida, 4111 Libra Drive, Orlando, FL, 32816*

Tong Wan¹, Ashley A. Geraets², Christopher A. Nix², Erin K. H. Saitta² and Jacquelyn J. Chini¹

¹*Department of Physics, University of Central Florida, 4111 Libra Drive, Orlando, FL, 32816*

²*Department of Chemistry, University of Central Florida, 4111 Libra Drive, Orlando, FL, 32816*

We investigated how changing the physical classroom impacted graduate teaching assistant (GTA) and student behaviors in tutorial sections of an introductory algebra-based physics sequence. Using a modified version of the Laboratory Observation Protocol for Undergraduate STEM (LOPUS), we conducted 35 observations over two semesters for seven GTAs who taught in different styles of classrooms (i.e., active learning classrooms and traditional classrooms). We found that both GTAs and students changed behaviors in response to a change from an active learning classroom to a traditional classroom. GTAs were found to be less interactive with student groups and to lecture at the whiteboard more frequently. Correspondingly, student behaviors changed as students asked fewer questions during one-on-one interactions. These findings suggest that the instructional capacity framework, which typically focuses on interactions between instructors, students and instructional materials, should also include interactions with the learning space. We suggest administrators and departments consider the impact of changing to a traditional classroom when implementing student-centered instruction and emphasize how to use classroom space in GTA professional development.

I. INTRODUCTION

To complement active learning instruction, active learning classrooms (ALCs) have been created as an alternative to some traditional classrooms (TCs). ALCs are technology-enriched learning spaces architecturally designed and furnished to support student-centered instruction [1,2], such as classrooms designed for SCALE-UP and Modeling Instruction [3,4]. Research in postsecondary education has suggested that ALCs have a positive impact on student learning outcomes (e.g., student performance) [5-10], but there is limited research about the impact of ALCs on student behaviors. For example, when observing undergraduate biology students in an ALC and a TC, Brooks (2012) found students to be significantly more “on-task” in the TC than in the ALC [5]. However, “on-task” was operationalized by behaviors typical of students in a TC setting (e.g., facing instructor and taking notes) [5]. Most studies have focused on the student perception of their behavior and found students to perceive increased motivation to attend and to participate in class when their course was held in an ALC rather than in a TC [8,9,11]. Therefore, the impact of the learning space on student behavior should be further investigated using classroom observations in postsecondary STEM courses.

In addition, research on learning spaces has suggested that the physical classroom has a possible impact on faculty instructor behavior [2]. For example, when an experienced biology instructor used the same instructional material in an ALC and a TC, the instructor was observed to lecture and to stand at the podium more frequently in the TC and to engage with student groups more frequently in the ALC [5,6]. In addition, Lasry, Charles and Whittaker (2014) studied the impact of learning space on postsecondary physics courses and found students performed the best on the Force Concept Inventory when their faculty instructor aligned their instructional style with the ALC [12]. Thus, if the classroom environment supports the instructor to lecture rather than engage with student groups, the instructional strategy might not have the anticipated impact on student outcomes.

Furthermore, the implementation of physics tutorial sessions has become a popular alternative to traditional recitations in postsecondary education since they provide a method for students to focus on their reasoning rather than getting the correct answer (e.g., *Tutorials in Introductory Physics* [13-16]). While literature supports a connection between graduate teaching assistant (GTA) and student behaviors [17], the literature about the impact of learning space on GTA behavior in tutorial sessions is almost non-existent. Therefore, to expand the literature about the impact of learning space on tutorial sessions led by GTAs, we explored the impact on interactions between GTA and students during introductory algebra-based physics tutorial sessions when the classroom was changed from an ALC to a TC. We used classroom observations to *directly* measure

student behavior and GTA instructional behavior. Specifically, we investigated the following research question: How did changing the designated learning space for introductory physics tutorials from an active learning classroom to a traditional classroom impact (1) GTA behavior and (2) student behavior?

II. METHODS

In this quasi-mixed method mono-strand conversion study [18], we quantized observation data to investigate if switching the classroom from an ALC to a TC impacted GTA and student behaviors.

A. Framework: instructional capacity including learning space

Our study builds on the instructional capacity (IC) framework by Cohen and Ball [19]. The original IC framework suggests interactions between the instructional material, instructor, and students influence instructional capacity [19]. For example, instructional materials inform GTA and student behaviors in the classroom. Meanwhile, instructors’ teaching beliefs and skills affect how instructors interact with students and instructional materials, and students’ attitudes and beliefs about learning impact their responses to instructors’ teaching and their engagement with instructional activities.

Since prior research suggests that the learning space impacts instructor and student behaviors, we argue that the IC framework should add a new element, learning space (e.g., furniture, technology, tools). This study allows us to evaluate the necessity of including learning space in the IC framework. We conducted classroom observations during which we observed both GTA and student behaviors in the same instructional context with the learning space varied.

B. Setting: physics “mini-studio” and learning space

The physics “mini-studio” combines tutorial and lab in three-hour weekly sessions for the two-semester introductory algebra-based physics sequence; the mini-studio is paired with a lecture section [20]. In this study, we focus on the tutorial portion, where students are expected to work in small groups on worksheets focused on student reasoning adapted from the *Tutorials in Physics Sensemaking* [21] and problem-solving skills. Typically, tutorial and lab were held in the same physical classroom: an ALC furnished with eight tables, each with four stools and a computer. Moveable whiteboards and markers were also available to each student group. Recently, a departmental decision was made to facilitate increased student enrollment by moving the tutorial portion out of the ALC and into a TC with 32 individual desks facing towards the front of the room. The room change occurred for the first-semester course (Physics I) in Fall 2019 and for the second-semester

course (Physics II) in Spring 2020. Since the room change only occurred with the tutorial portion, this study focused only on the tutorial. In addition to the classroom switch, the time allotted for tutorial was reduced from 75 minutes (in ALC) to 50 minutes (in TC) [22].

C. Participants: “mini-studio” GTAs

As an attempt to control for variation in classroom practice across GTAs and courses they teach, we limited our investigation to seven (out of thirteen) GTAs who taught the same course in both semesters. Three of the seven GTAs led mini-studio sections for Physics I and the other four for Physics II. Four GTAs were new to teaching mini-studio in Fall 2019, and three GTAs taught mini-studio sections in a previous semester. For both courses, GTAs attended weekly meetings to discuss the instructional material and possible student ideas, and to reflect on their teaching in previous weeks. GTAs also participated in three to four practice teaching sessions with a mixed-reality classroom simulator [23] each semester to practice implementing teaching strategies suggested to increase student participation.

D. Classroom observations

Four researchers (C.M.D., T.W., A.A.G. and C.A.N.) observed the GTAs teach their mini-studio sections two to three times each semester for a total of 35 observations. A modified version of the Laboratory Observation Protocol for Undergraduate STEM (LOPUS) was used to code GTA behaviors (e.g., GTA posing question to whole class) and student behaviors (e.g., working on tutorial worksheets) in 2-minute intervals [24,25]. GTA and student codes are displayed in Table I. Codes are not mutually exclusive (i.e., it is possible more than one behavior could occur in the same 2-minute interval). As part of another project [26], we conducted observations using the same protocol prior to this study. To investigate inter-rater reliability (IRR), researchers observed the same sections in groups of two or three both prior to and during this study. We calculated Gwet’s AC1 for each code as a metric for IRR [27]. IRR metrics provide evidence that our observation data was reliably collected. During group observations in the Fall 2019 semester, IRR values for each code during tutorial for pairs of observers ranged between 0.679 and 1. A Gwet’s AC1 value greater than 0.81 can be interpreted as near perfect agreement, and a value between 0.6 and 0.8 can be interpreted as moderate agreement [28].

E. Data analysis

We analyzed observation data from Physics II to investigate the impact of changing learning space on GTA and student behaviors. We note that the GTAs participated in professional development activities in both semesters and gained more teaching experience. To tease out the impact of professional development as well as GTAs gaining teaching

experience, we also analyzed data from Physics I where the tutorial occurred in the TC both semesters.

We calculated the percentage of 2-minute intervals each code occurred for every observation. For example, if the duration of tutorial was 50 minutes (25 2-minute intervals) and Lecture was coded for 10 2-minute intervals, then the percentage of 2-minute intervals for Lecture would be 40%.

TABLE I. GTA and student behavior codes [29]

Behavior Category	Behavior	Description
Typical (GTA)	Lecture (Lec)	Talking at whole class
	Demo/Video (D/V)	Showing demo or video
	Following-Up (Fol)	Following up or providing feedback on completed activity
	Real-Time Writing (RtW)	Writing on board, doc cam, etc. for whole class
	Monitoring (M)	Available to students but not currently engaging students
Interactive (GTA)	TA-Initiates (TI)	Initiating interaction with small group or individual
	One-on-One Talk (1o1-Talk)	Engaging with small group or individual
	One-on-One TA Posing Question (1o1-TPQ)	Posing question to small group or individual
	Posing Question (PQ)	Posing question to whole class
	Verbal Feedback (VF)	Providing verbal feedback about student idea
Non-instructional (GTA)	Verbal Monitoring (VM)	Verbally checking in on small group or individual
	Administration (Adm)	Performing an administrative task
	Waiting (W)	Unavailable to students
	Student-Initiates (SI)	Initiating interaction with GTA
	Student Question (SQ)	Asking a question in front of whole class
Interactive (Student)	One-on-One Student Question (1o1-SQ)	Asking a question in front of small group or individually
	Worksheet (Wks)	Working on the tutorial in their small groups

Then, we compared the percentage of 2-minute intervals for each code between Fall 2019 and Spring 2020 for GTAs in each course. Due to small sample size and uneven number of observations per semester, we used the Wilcoxon rank-sum test [30]. The analysis was conducted in Rstudio using the function “wilcox.test” [31]. We used the “wilcoxonr” function in the rcompanion package [32] to obtain a modified correlation coefficient (r) commonly paired with Wilcoxon tests to investigate effect size [33,34]. Modified r can be interpreted similarly to Cohen’s d (large: $d > 0.5$, medium: $d > 0.3$, small: $d > 0.1$) [33,35]. Also, due to small sample size and a medium or large effect size, we discuss codes with $p < 0.1$ and $p < 0.05$.

III. RESULTS

A. Changes in GTA and student behaviors for GTAs who taught in the TC both semesters

We compared GTA and student behaviors for Physics I observations. We did not find significant differences in the percentages for student behaviors as described by the modified LOPUS; however, three out of twelve GTA behaviors were statistically different, as shown in Figure 1. In the spring semester, we found the percentage for Following-Up (Fol, $p = 0.057$) and Waiting (W, $p = 0.036$) increased with large effect sizes ($r = 0.514$ and $r = 0.555$, respectively) in comparison to fall semester observations. Also, the percentage for Posing Questions in front of the whole class (PQ, $p = 0.082$) decreased with a medium effect size ($r = -0.485$) during the spring semester. The results suggest that GTAs became somewhat less interactive when they continued to teach in the TC, which is consistent with previous findings that experienced GTAs tend to become less interactive when compared to new GTAs [31].

B. Impact of changing classrooms on GTA behaviors

We found that seven GTA behaviors changed for GTAs who switched from teaching in the ALC to the TC in the spring semester, as shown in Figure 2. The percentage for Lecture (Lec, $p = 0.047$) and Monitoring (M, $p = 0.074$) increased with medium effect sizes ($r = 0.449$ and $r = 0.407$, respectively) and the percentage for Real-Time Writing (RtW, $p = 0.0170$) increased with a large effect size ($r = 0.543$) when GTAs switched to teaching in the TC from the ALC. We did not find the percentages for Lec, RtW or M to change for GTAs who taught both semesters in the TC when comparing fall and spring semester data. These results suggest the switch from the ALC to the TC likely impacted GTA behaviors by promoting behaviors more typical to TCs.

Additionally, the percentage for TA-Initiates (TI, $p = 0.037$) and Waiting (W, $p = 0.0542$) decreased with medium effect sizes ($r = -0.476$, $r = -0.438$, respectively) and the percentage for One-on-One Talk (1o1-Talk, $p = 0.010$.) and One-on-One TA Posing Question (1o1-TPQ, $p = 0.015$)

decreased with large effect sizes ($r = -0.570$ and $r = -0.552$, respectively) in the spring semester when compared to the fall semester. As mentioned previously, the percentages for TI, 1o1-Talk and 1o1-TPQ did not change across semesters for GTAs who taught in the TC both semesters. Therefore, decreased percentages of TI, 1o1-Talk and 1o1-TPQ were likely due to the change in learning space. However, the percentage for W increased for GTAs who taught in the TC both semesters. It is unclear why this GTA behavior displayed opposite effects under these different conditions. It may be due to the small sample size.

In summary, we found interactive GTA behaviors (e.g., TI, 1o1-Talk) to decrease and typical GTA behaviors (e.g., Lec, RtW) to increase for GTAs who switched from teaching in the ALC to the TC. These results suggest changing the physical classroom had an impact on GTA behavior in which GTAs became less interactive in the TC. This finding is consistent with the results of Brooks’ study about an experienced biology faculty instructor [5].

C. Impact of changing classrooms on student behaviors

As shown in Figure 2, the percentage for Student Questions asked in front of the whole class (SQ, $p = 0.038$) increased and the percentage for One-on-One Student Question (1o1-SQ, $p = 0.047$) decreased with medium effect sizes ($r = 0.474$ and $r = -0.449$, respectively) when the classroom was changed from an ALC to a TC. In comparison, student behavior in tutorial sections that remained in the TC both semesters did not change as described by the modified LOPUS. These results suggest the switch from the ALC to the TC impacted student behaviors, with students becoming less interactive with their GTA and exhibiting more typical student behavior in a TC. This result is expected since we found GTAs altered their behavior to align with the new classroom setting.

FIGURE 1. Percentage of 2-minute intervals for GTA behaviors that were significantly different without room change. * $p < 0.1$, ** $p < 0.05$

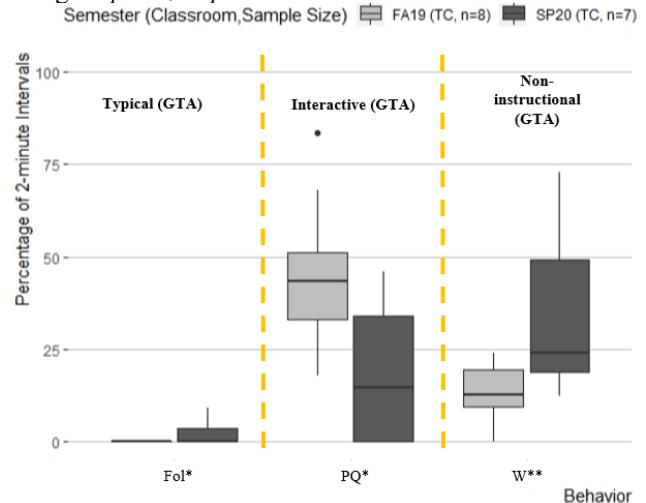
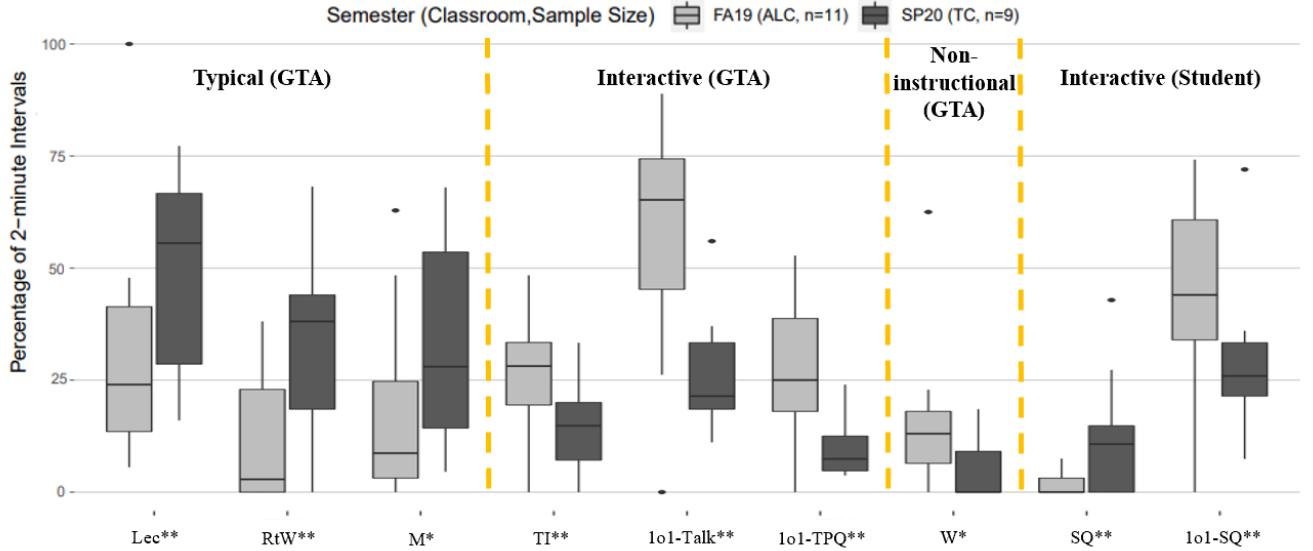


Figure 2. Percentage of 2-minute intervals for GTA and student behaviors that were significantly different with room change. * $p < 0.1$, ** $p < 0.05$



IV. DISCUSSION AND IMPLICATIONS

In our study, we found interactive GTA behaviors decreased and GTA behaviors typical of a TC increased when the designated classroom for tutorial switched from an ALC to a TC. In addition, student behaviors changed similarly to GTA behaviors in response to the classroom switch. Since we found an impact of learning space on both GTA and student behaviors, we suggest modifying the IC framework to include learning space as a fourth element, especially since ALCs have been designed to support the implementation of student-centered instructional material.

As an effort to mitigate the potential negative impact on student learning, GTAs were instructed during prep meetings to rearrange the individual desks into clusters in the TC. However, administrators did not approve of a permanent change to the layout of the TC, which meant GTAs would have to rearrange the desks at the beginning and end of class. We also observed a variation in effort across GTAs to rearrange the desks in the TC; some GTAs rearranged desks into clusters while some GTAs left the desks in rows. The configuration of the desks in the TC could have impacted the interaction between GTA and students. If GTAs left the desks in rows, the focal point becomes the front of the room at the whiteboard [36]. On the contrary, if GTAs rearranged the desks in clusters or students sat at tables, students could focus on each other instead of the GTA [36]. Therefore, the arrangement of the desks in rows could have influenced GTAs to lecture more frequently in the TC while the tables in the ALC promoted GTAs to interact with students more frequently in the small group setting.

Consistent with the IC framework, there exists a two-way interplay between GTA and student behaviors in response to

the change in learning space. During the prep meetings, several GTAs voiced their concern about student resistance when they directed students to make groups in the TC. This type of student resistance, referred to as “ignoring the teacher”, is not uncommon in active learning environments [37], and GTAs might not be equipped with strategies to overcome such difficulties. Furthermore, it is possible the student resistance stemmed from a violation of their expectation of how their GTA should teach in a TC [38]. In turn, GTAs reacted to the student resistance by increasing time spent lecturing because they believe their students prefer that type of teaching method [20]. Therefore, we suggest STEM departments and administrators to consider how changing the learning space might affect GTA and student behavior, especially since postsecondary science faculty instructors have reported that ALCs more easily support the implementation of student-centered activities [39].

To be clear, our findings do not imply TCs can never be used for non-traditional recitations. For example, clusters of individual desks have been reported to have the same impact on student learning outcomes (e.g., exam scores) as the use of tables with student-centered instruction [40]. Instead, we suggest departments provide GTA professional development to inform GTAs about how to use the classroom to support the instructional material, especially if a tutorial-style recitation is going to be led by GTAs in a TC.

ACKNOWLEDGMENTS

This work is supported in part by the National Science Foundation DUE Grant No. 1725554. We thank all GTAs who participated in our study.

[1] R. S. Grabinger and J. C. Dunlap, Rich environments for active learning: A definition. [ALT-J, 3\(2\), 5-34 \(1995\)](#).

[2] R. Talbert and A. Mor-Avi, A space for learning: An analysis of research on active learning spaces. [Heliyon, 5\(12\) \(2019\)](#).

[3] E. Brewe, V. Sawtelle, L. H. Kramer, G. E. O'Brien, I. Rodriguez and P. Pamelá, Toward equity through participation in Modeling Instruction in introductory university physics. [Physical Review Special Topics-Physics Education Research, 6\(1\), 010106 \(2010\)](#).

[4] R. J. Beichner, J. M. Saul, D. S. Abbott, J. J. Morse, D. L. Deardorff, R. J. Allain, S. W. Bonham, M. H. Dancy and J. S. Risley, (2007). The student-centered activities for large enrollment undergraduate programs (SCALE-UP) project. Research-based reform of university physics, [1\(1\), 2-39 \(2007\)](#).

[5] D. C. Brooks, Space and consequences: The impact of different formal learning spaces on instructor and student behavior. [Journal of Learning Spaces, 1\(2\), n2 \(2012\)](#).

[6] S. Cotner, J. Loper, J. D. Walker, and D.C. Brooks, " It's Not You, It's the Room"—Are the High-Tech, Active Learning Classrooms Worth It?. [Journal of College Science Teaching, 42\(6\), 82-88 \(2013\)](#).

[7] Y. Shi, H. Yang, J. MacLeod, J. Zhang and H. H. Yang, College Students' Cognitive Learning Outcomes in Technology-Enabled Active Learning Environments: A Meta-Analysis of the Empirical Literature. [Journal of Educational Computing Research, 0735633119881477 \(2019\)](#).

[8] R. F. Yuretich, and L. C. Kanner, Examining the effectiveness of team-based learning (TBL) in different classroom settings. [Journal of Geoscience education, 63\(2\), 147-156 \(2015\)](#).

[9] E. L. Park, and B. K. Choi, Transformation of classroom spaces: Traditional versus active learning classroom in colleges. [Higher Education, 68\(5\), 749-771 \(2014\)](#).

[10] P. H. P. Chiu and S. H. Cheng, Effects of active learning classrooms on student learning: a two-year empirical investigation on student perceptions and academic performance. [Higher Education Research & Development, 36\(2\), 269-279 \(2017\)](#).

[11] J. D. Walker, D. C. Brooks, and P. Baehler, Pedagogy and space: Empirical research on new learning environments. [Educuse Quarterly, 34\(4\), n4 \(2011\)](#).

[12] N. Lasry, E. Charles, and C. Whittaker, When teacher-centered instructors are assigned to student-centered classrooms. [Physical Review Special Topics-Physics Education Research, 10\(1\), 010116 \(2014\)](#).

[13] L. C. McDermott, P. S. Shaffer, and M. D. Somers, Research as a guide for teaching introductory mechanics: An illustration in the context of the Atwood's machine. [American Journal of Physics, 62\(1\), 46-55 \(1994\)](#).

[14] N. D. Finkelstein and S. J. Pollock, Replicating and understanding successful innovations: Implementing tutorials in introductory physics. [Physical Review Special Topics-Physics Education Research, 1\(1\), 010101 \(2005\)](#).

[15] R. M. Goertzen, E. Brewe, L. H. Kramer, L. Wells and D. Jones, Moving toward change: Institutionalizing reform through implementation of the Learning Assistant model and Open Source Tutorials. [Physical Review Special Topics-Physics Education Research, 7\(2\), 020105 \(2011\)](#).

[16] M.S. Sabella, Implementing tutorials in introductory physics at an inner-city university in Chicago, [In Physics Education Research Conference Proceedings \(pp. 79-82\) \(2002\)](#).

[17] J. B. Stang and I. Roll, Interactions between teaching assistants and students boost engagement in physics labs. [Physical Review Special Topics-Physics Education Research, 10\(2\), 020117 \(2014\)](#).

[18] C. Teddlie and A. Tashakkori, Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences. (Sage, Los Angeles, 2009) Vol.1 p. 145 – 151.

[19] D. K. Cohen and D. L. Ball, Instruction, Capacity, and Improvement; [CPRE Research Report Series RR-43; Consortium for Policy Research in Education: University of Pennsylvania, Graduate School of Education, 1999](#).

[20] M. Wilcox, Y. Yang, and J. J. Chini, Quicker method for assessing influences on teaching assistant buy-in and practices in reformed courses. [Physical Review Physics Education Research, 12\(2\), 020123 \(2016\)](#).

[21] A. Elby, R. E. Scherr, R. M. Goertzen and L. Conlin, Open Source Tutorials in Physics Sensemaking (2008).

[22] We recognize the change in time allotted for recitation as a limitation to our study. We did not use a method to disentangle the influence of the change in duration from the change in learning space.

[23] L. A. Dieker, C. L. Straub, C. E. Hughes, M. C. Hynes and S. Hardin, Learning from virtual students. [Educational Leadership, 71\(8\), 54-58 \(2014\)](#).

[24] J. B. Velasco, A. Knedeisen, D. Xue, T. L. Vickrey, M. Abebe and M. Stains, Characterizing instructional practices in the laboratory: The laboratory observation protocol for undergraduate STEM. [Journal of Chemical Education, 93\(7\), 1191-1203 \(2016\)](#).

[25] It may be surprising that we used a laboratory observation protocol to observe recitations. This is to some extent an artifact of the change investigated in this study: initially, the tutorial and laboratory were integrated, and we used LOPUS to observe the full mini-studio session. We continued using LOPUS once the room changed the tutorial portion to allow direct comparison. The behaviors described by LOPUS encompass behaviors expected in both the tutorial and lab portions of the mini-studio, as informed by the course designers and trainers (authors of this proceedings).

[26] T. Wan, A. A. Geraets, C. Doty, E. K. H. Saitta and J. J. Chini, Characterizing science graduate teaching assistants' instructional practices in reformed laboratories and tutorials. [International Journal of STEM Education, 7\(1\), 1-21 \(2020\)](#).

[27] K. Gwet, Kappa statistic is not satisfactory for assessing the extent of agreement between raters, [Statistical Methods for Inter-Rater Reliability Assessment Series 2, 9 \(2002\)](#).

[28] K. L. Gwet, Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters 4th ed. (Advanced Analytics, Gaithersburg, 2014), Vol. 1 p. 104 - 107.

[29] We did not include the whole class discussion code because whole class discussion did not occur during observations. We also did not include the code for students taking quiz/exam because the students are supposed to take the quiz at the beginning of the lab portion, not in during the recitation portion.

[30] P. D. Bridge and S. S. Sawilowsky, Increasing physicians' awareness of the impact of statistics on research outcomes: comparative power of the t-test and Wilcoxon rank-sum test in

small samples applied research. *Journal of clinical epidemiology*, **52**(3), 229-235 (1999).

[31] RStudio Team (2020). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. <http://www.rstudio.com/>

[32] Salvatore Mangiafico (2020). rcompanion: Functions to Support Extension Education Program Evaluation. R package version 2.3.25. <https://CRAN.R-project.org/package=rcompanion>

[33] C.O. Fritz, P.E. Morris, and J.J. Richler, Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, **141**(1), 2-18 (2012).

[34] M. Tomczak and E. Tomczak, The need to report effect size estimates revisited. An overview of some recommended measures of effect size, *Trends in Sport Sciences*, **1**(21), pp. 19-25 (2014).

[35] J. Cohen, Statistical power analysis for the behavioral sciences 2nd ed. (Erlbaum, Hillsdale, NJ, 1988).

[36] C. Smith, The Influence of Hierarchy and Layout Geometry in the Design of Learning Spaces. *Journal of Learning Spaces*, **6**(3), 59-67 (2017).

[37] S. B. Seidel and K. D. Tanner, "What if students revolt?"—considering student resistance: origins, options, and opportunities for investigation. *CBE—Life Sciences Education*, **12**(4), 586-595 (2013).

[38] J. D. Gaffney, A. L. H. Gaffney, and R. J. Beichner, Do they see it coming? Using expectancy violation to gauge the success of pedagogical reforms. *Physical Review Special Topics-Physics Education Research*, **6**(1), 010102 (2010).

[39] J. Robert, C. Ramsay, S. E. Ades, K. C. Keller and C. Palma, Active Learning Spaces: Matching Science Classrooms with Pedagogy. In Active Learning in College Science (Springer, Cham, 2020), Vol. 1 pp. 483-498.

[40] R. S. Muthyalu and W. Wei, Does space matter? Impact of classroom space on student learning in an organic-first curriculum. *Journal of Chemical Education*, **90**(1), 45-50 (2013).