ARTICLE
Class Size and Student Performance in a Team-Based Learning Course

Minna Ng[1] and Thomas M. Newpher[1,2]
[1]Department of Psychology and Neuroscience, Duke University; [2]Duke Institute for Brain Sciences, Duke University.
Address correspondence to: Dr. Thomas Newpher, Department of Psychology and Neuroscience, 308 Research Drive, Duke University, Durham, NC 27708.

High-enrollment university courses can be associated with decreased student learning and course satisfaction. In these large classes, students report feelings of isolation, reduced faculty interaction, and less motivation. Here we address whether team-based learning (TBL), a highly interactive and collaborative form of active learning, can improve the student experience in larger undergraduate neuroscience courses. Specifically, we analyzed student performance on summative assessments, as well as survey responses on measures of the classroom environment from a single TBL course, taught over a range of enrollment sizes (19-103 students). While the higher enrollment course terms had decreased ratings of course quality compared to the lower enrollment terms, we also found that student performance on exams was similar across all course term sizes. Furthermore, we observed no differences across class sizes for most measures of classroom dynamics and course characteristics. Taken together, our data suggest that the content knowledge outcomes and many aspects of the classroom environment were not negatively impacted in the higher enrollment versions of this TBL course.

Key words: Team-Based Learning (TBL); Class Size; Active Learning; Undergraduate Neuroscience Education; STEM

Class size is one major factor thought to impact student learning, achievement, and classroom dynamics, as well as course and instructor ratings (Cash et al., 2017; Wright et al., 2019). In higher education, large class sizes are associated with student feelings of isolation and anonymity, both of which can lead to decreased student motivation and retention (Carbone and Greenberg, 1998; Mulryan-Kyne, 2010; Wulff et al., 1987). In addition, instructors report difficulty engaging with students and reduced faculty-student interactions in large classes (Carbone and Greenberg, 1998; Gibbs et al., 1996; Mulryan-Kyne, 2010; Wulff et al., 1987). Furthermore, small class sizes are associated with increased student engagement (Gleason, 2012), as well as higher student ratings of the course and instructor quality (Bedard and Kuhn, 2008; Benton and Cashin, 2012; Kwan, 1999; Mandel and Sussmuth, 2011; Monks and Schmidt, 2011; Sapelli and Illanes, 2016; Westerlund, 2008). There are a number of possible explanations for the positive classroom dynamics observed with smaller classes. These could include stronger student-faculty relations, greater personal attention, and more frequent use of active learning and interactive pedagogies (Cash et al., 2017; Wright et al., 2019).

Not surprisingly, students in smaller classes typically self-report greater learning (Benton and Pallett, 2013; Monks and Schmidt, 2011). However, direct measures of student learning related to class size are quite variable. For example, some studies report mixed results or no effect related to class size (Bellante, 1972; De Paola et al., 2013; Edgell, 1981; Gleason, 2012; Hancock, 1996; Hill, 1998; Jarivs, 2007; Kennedy and Siegfried, 1997; Matta et al., 2015; Olson et al., 2011; Raimondo et al., 1990), while others report increased student learning and performance with small class sizes (Arias and Walker, 2004; Baeten et al., 2010; Bandiera et al., 2010; Becker and Powers, 2001; Freeman et al., 2014; Gibbs et al., 1996; Kogl Camfield et al., 2016; Kokkelenberg et al., 2006; Westerlund, 2008). To date, the relationship between class size and student learning remains a highly debated topic in higher education (Ake-Little et al., 2020).

There are several possible explanations for the variable findings between studies on class size and learning. These could include the demographics of the student populations, the types of schools, and the teaching methods employed (Ballen et al., 2018). Indeed, it remains unknown which types of teaching approaches are best suited for student learning in large university classrooms. One might predict that more engaging and structured teaching methods that motivate students, encourage them to attend, and provide opportunities to practice would help improve student learning in larger university courses (Eddy and Hogan, 2014). Consistent with this idea, a meta-analysis comparing the effectiveness of active learning pedagogies to lecture-based courses found that active learning significantly increases student achievement in all class sizes (0-50, 51-110, and >110) (Freeman et al., 2014). However, the study also found that courses with fewer than 50 students showed greater learning gains than groups of 51 or more students (Freeman et al., 2014), suggesting that even mixed active learning approaches work better in lower enrollment courses. These findings highlight the need to better understand the exact types of active learning methods and course structures that improve student learning in high enrollment courses.

One active learning method that has the potential to work well for large university courses is team-based learning (TBL). In TBL, students work together in permanent teams throughout the semester. Not only do students complete individual readiness assurance tests (iRATs) during each learning module, but they also

collaborate with peers on team readiness assurance tests (tRATs) and application activities, designed to promote learning on Bloom's lower-order (recall, understand) and higher-order (apply, analyze, evaluate, synthesize) cognitive outcomes (Bloom, 1956; Michaelsen et al., 2004; Michaelsen and Sweet, 2008). Of note, TBL was created over 40 years ago for the purpose of transforming large lecture halls into inclusive, interactive spaces that provide strong motivation to attend, participate, and learn (Michaelsen et al., 1982). Importantly, many of the design elements thought to promote learning in TBL (immediate feedback, retrieval practice, and distributed practice) would be present to the same extent in both small and large TBL courses (Butler et al., 2014; Dunlosky et al., 2013; Schmidt et al., 2019). While some evidence indicates that larger TBL courses are harder for instructors to manage (Allen et al., 2013; Thompson et al., 2007a; Thompson et al., 2007b), to date, the relationship between class size and student performance in TBL has not been formally addressed.

Given the highly structured and interactive design of TBL courses, we predicted that student performance on exam questions that tested lower- and higher-order learning outcomes would be similar across a range of TBL course term sizes. Furthermore, we expected that measures addressing the classroom environment would not decline as enrollment size increased. To this end, we analyzed student exam performance and end-of-semester course evaluations from a single undergraduate neuroscience course. This single TBL course was taught across a range of enrollment sizes (19-103) over four course terms. We found that the lower enrollment versions of the course had higher student ratings of course quality than the higher enrollment versions, even though all course terms were taught by the same instructor. However, exam performance and several measures of the classroom environment did not differ across a range of course term sizes, suggesting that for this TBL course, many aspects of the learning experience were not negatively impacted in higher enrollment course terms.

MATERIALS AND METHODS

This study was approved by the Duke University Institutional Review Board. The course was taught at a private university in the southeastern United States. For ratings of the classroom environment, we used data from end-of-semester course evaluation surveys from four different terms of the same undergraduate neuroscience course. All four course terms were taught by the same instructor. Course evaluations were collected and provided by the Office of Assessment and made available to the instructors and researchers after grades were posted by the Registrar's Office. Students enrolled in this in-person undergraduate neuroscience course submitted anonymous campus-wide course evaluations on their perceptions of a range of questions addressing classroom dynamics and course characteristics. Twenty minutes of in-class time were reserved to complete the voluntary course evaluations, although students also had the option to take evaluations outside of class time. Seven different

measures of the classroom environment were analyzed, except for Term 1 and Term 2 where "course difficulty" data were not collected. The students were asked to provide responses to the following questions on their end of semester course evaluations:

1. The course had clearly defined student learning objectives.
2. The course had clear expectations for assignments and other work.
3. The course had a welcoming and inclusive classroom environment.
4. The course helped me to effectively communicate ideas orally.

(For questions 1-4, response options were on a Likert rating scale: 5 = strongly agree, 4 = agree, 3 = neutral, 2 = disagree, and 1 = strongly disagree; or NA = not applicable.)

5. Please characterize the difficulty of the subject matter.

(Response options were on a Likert rating scale: 5 = very high, 4 = high, 3 = moderate, 2 = low, and 1 = very low; or NA = not applicable.)

6. Give an overall rating for the quality of this course.
7. Give an overall rating for the quality of instruction.

For questions 6-7, response options were on a Likert rating scale: 5 = excellent, 4 = very good, 3 = average, 2 = marginal, and 1 = poor; or NA = not applicable.

To measure Bloom's cognitive outcomes, we analyzed average student performance on summative exam questions taken during each of the four course terms. The summative exam questions tested students on all content learned during the semester, including Bloom's lower-order and higher-order cognitive outcomes. As summative exam questions were slightly different between course terms, we restricted our analysis to questions that were identical and reused between the corresponding semesters, Term 1-Term 2 and Term 3-Term 4. The summative exam questions came from two different midterm exams. One midterm was offered at the halfway point of the course term and the second at the end of the course term. Exam questions were multiple choice and covered the range of Bloom's taxonomic levels (Bloom, 1956). Student performance was tracked across all identical test questions between terms, referred to as "All Questions" (Tables 3, 4, and 5). The exam questions were then divided into two categories based on Bloom's taxonomy: lower-order cognitive outcomes for recall and understanding level questions (Lower Bloom's), and higher-order cognitive outcomes for application, analysis, and evaluation (Higher Bloom's). See Appendix 1 for examples of lower- and higher-order questions. The number of student exams used to measure exam performance was not identical to the starting enrollment for each course term. In some cases, students withdrew from the course and did not complete one or both of the midterms. While the large majority of students completed their midterms online using the course management software (Sakai ©), several students completed midterms on paper and their records were not stored. Performance on assessments and responses on end-of-semester course evaluations were

Table 1. Characteristics of the 4 TBL Course Terms

Term	Enrollment	TBL Modules	Meetings	Discussion	Term Length	Exams	Team Size
Term 1	19	8	27	No	6 (summer)	2	6
Term 2	103	8	28	No	14 (fall)	2	6
Term 3	101	12	42	Yes	14 (fall)	2	6
Term 4	64	10	40	Yes	14 (fall)	2	6

Table 1. Characteristics of the 4 TBL Course Terms. Enrollment = number of students enrolled at the start of the semester. TBL Modules = number of TBL modules or units during the semester (each module consisted of 1 pre-lecture, 1 readiness assurance, and 1 application activity). Meeting = number of class meetings during the semester; Discussion = indicates that the course term had a weekly discussion section. Term length = length of the course term in weeks. Exams = number of summative exams. Team Size = size of small learning groups.

analyzed using non-parametric Mann Whitney U tests, performed in IBM SPSS®.

Precise definitions and ranges for small, medium, and large university classes vary between studies (Cash et al., 2017; Freeman et al., 2014; Wright et al., 2019). As such, we did not label these classes as small, medium, or large. Rather, we refer to them by their enrollment size. In all classes, tables shared among teammates were the same size and style and were spaced apart similarly. The courses in these analyses used the same classroom space, except Term 1, which was held in a smaller classroom with seating for 20 students. Classrooms for all terms and sizes were single-level, as opposed to stadium seating.

The same male instructor (T.M.N.) taught this course and had six years of teaching experience (one year with TBL) prior to the start of Term 1. The TBL course analyzed in this study was a 200-level intermediate undergraduate neuroscience course and, at the time, was a graduation requirement for the Neuroscience major. The following is the course description that appeared on the syllabus, and below are the six learning objectives for the course.

Course Description:
"In this course, learners explore the organization of neural systems that allow us to sense our environment, plan and execute complex movements, encode and retrieve memories, and experience a wide range of emotions. We also examine the development of the brain and spinal cord and how changes in the structure and function of these neural systems underlie the devastating effects of neurological and psychiatric disorders."

Course Learning Objectives:
-Describe how neurons generate, propagate, and communicate electrical signals.
-Recall the major steps of synaptic transmission and the signaling pathways that drive synaptic plasticity.
-Compare and contrast sensory pathways and describe how stimuli generate electrical signals in sensory neurons to produce different sensations.
-Characterize the pathways and brain regions that control movement.
-Explain how cortical and subcortical brain structures develop and control cognitive processes, including memory, emotion, attention, and planning.
-Predict how lesions in neural systems could lead to the development of neurological and psychiatric disorders.

The characteristics for all four course terms are shown in Table 1. While all four terms covered similar neuroscience content, there were some notable differences between course terms. First, Term 1 was a summer session course, while Terms 2, 3, and 4 where all taught during the fall semester. In addition, Terms 3 and 4 included discussion section meetings and allowed for more TBL modules compared to Terms 1 and 2, which did not have discussion section meetings. This meant that Terms 3 and 4 had more class meetings and TBL modules, allowing for more material to be covered in a term. To account for this, pairwise comparisons were made only between corresponding terms: Term 1 with Term 2, and Term 3 with Term 4.

The standard TBL unit or module consists of two major phases: the readiness assurance process (RA) and the application activity (Michaelsen et al., 2004). The RA process starts with students reading basic background material or watching videos outside of class, followed by individual and team readiness assurance tests (iRATs and tRATs) at the start of the next class period. The iRATs and tRATs serve as formative assessments and are designed to promote student preparedness for the in-class team activities that occur throughout the module. Next, in the application phase of the module, students deepen their learning by collaborating with teammates on complex problems which help them achieve the higher-order cognitive outcomes (Michaelsen et al., 2004). For all four terms, the course was taught using a standard TBL format (Haidet et al., 2014), except that a pre-lecture was included in the class period prior to the readiness assurance process- described in (Ng and Newpher, 2020). The readiness assurance process (iRAT, tRAT, feedback and mini-lecture) was completed during the second class-period of each TBL module. The application activities were completed on the third day of each module.

Teams were formed based on the amount of experience students had in the field of neuroscience. Specifically, each team had a mix of students with none, some, or a great deal of neuroscience coursework and research

Table 2. Student Demographics of the 4 TBL Course Terms

Term	W	Major	Nonmajor	Undeclared	4th	3rd	2nd	1st	Grad	Visiting
Term 1	15.8	15.8	15.8	26.3	10.5	21.1	26.3	0.0	10.5	31.6
Term 2	2.0	65.2	14.5	18.4	28.2	47.6	22.3	0.0	1.9	0.0
Term 3	3.9	65.3	13.9	20.8	19.8	51.4	28.9	0.0	0.0	0.0
Term 4	1.5	64.1	25.0	9.4	25.0	60.9	10.9	1.6	0.0	1.6

Table 2. Demographics of the 4 TBL Course Terms. W = percentage of students withdrawing from the course. Major = percentage of students majoring in neuroscience. Nonmajor = percentage of students majoring in fields outside of neuroscience. Undeclared = percentage of students that have not yet declared a major. 4th year = percentage of students in their 4th year of study (or rising 4th year). 3rd year = percentage of students in their 3rd year of study (or rising 3rd year). 2nd year = percentage of students in their 2nd year of study (or rising 2nd year). Graduate = percentage of students enrolled in graduate school or post-baccalaureate programs. Visiting = percentage of students from outside universities/colleges.

experience. Teams were permanent and worked together for the entire semester. On average, the TBL learning teams each had six members. Although, in a few cases, teams had five or seven members if there were uneven numbers of students in the class. Evaluations of teammates were submitted halfway through the term and again at the end of the term.

Students enrolled in this course typically include third and fourth-year students majoring in neuroscience. Other typical students include those who are majoring in related disciplines such as biology and psychology, second-year students who have yet to declare a major, as well as visiting students and graduate/post-baccalaureate students (Table 2). The summer term course in this study (Term 1) met five days per week for 75-minute class sessions over six weeks. The fall term courses met two times per week for Term 2 (75-minute class sessions) and three times per week for Terms 3 and 4 (two 75-minute sessions and one 50-minute session) over 14 weeks. All fall term classes were taught in the early afternoon (1:25pm start) and the summer term course was taught late morning (11:00am start).

RESULTS
We first looked for differences between Term 4 (64 students) and its corresponding higher enrollment version, Term 3 (101 students) (1.58-fold size difference). Compared to Term 3, students in Term 4 gave significantly higher ratings for only one measure related to the classroom environment: course quality (Table 3). In addition, there were no significant differences for summative exam performance between students in Terms 3 and 4.

Next, we compared Term 1 (19 students) to its corresponding higher enrollment term (Term 2 -103 students) (5.4-fold size difference). When compared to Term 2, students in Term 1 gave significantly higher ratings for these measures related to the classroom environment: course quality, instructor quality, and oral communication skills (Table 4). Summative assessment performance and other course characteristics, however, were not significantly different between students in Terms 1 and 2.

Term 1 was taught during a 6-week summer session and contained a greater percentage of visiting and graduate/post-baccalaureate students than Term 2 (Table 2). We considered the possibility that these students had different levels of background knowledge that may have impacted their performance on exams. To account for this, we performed an analysis removing the visiting and graduate/post-baccalaureate students from both terms. When only considering majors, nonmajors, and undeclared undergraduate students in our analysis, we still did not find significant differences in summative exam performance between the low and high enrollment course terms (Table 5).

DISCUSSION
Summary
Here we have measured exam performance and student self-report of several measures of the classroom environment in a single TBL-taught course. Our findings demonstrate that, for this particular course, student performance on exam questions across Bloom's taxonomy and student self-report of a welcoming and inclusive classroom environment were not negatively impacted in higher enrollment course terms. Importantly, these findings suggest that, for this single course analyzed, the TBL approach was able to scale to the larger classroom setting. Below we discuss the significance of these findings and propose strategies to improve the experience for students in high enrollment university courses.

Class Size and the Classroom Environment in TBL
Regardless of class size, students in a TBL class have a small-group learning experience for the entire duration of the semester. Discussions with teammates and frequent interactions with members of other groups, would be predicted to help maintain a social, engaging, and positive classroom environment in large TBL courses. Consistently, we saw no differences across class size comparisons for ratings of "a welcoming and inclusive classroom environment" (Tables 3 and 4). Furthermore, in our comparison between Term 4 (64 students) and Term 3 (101 students), only one of the classroom environment measures was rated significantly greater by students in Term 4: course quality (Table 3). No significant differences were found in the other measures: "clearly

Table 3. Student Ratings and Exam Scores for Terms 3 and 4

	64 Students (Term 4)			101 Students (Term 3)				
	n	mean	SD	n	mean	SD	(U)	p value
Classroom Environment:								
Learning Objectives	37	4.81	0.74	69	4.67	0.87	(1129.00)	0.125
Clear Expectations	37	4.76	0.76	69	4.61	0.81	(1106.50)	0.122
Course Difficulty	37	3.76	0.64	69	3.70	0.73	(1220.00)	0.681
Oral Communication	37	3.83	0.95	69	3.80	1.02	(1046.00)	0.974
Welcoming/Inclusive	37	4.68	0.78	69	4.62	0.89	(1255.50)	0.849
Course Quality	**37**	**4.62**	**0.55**	**68**	**4.35**	**0.66**	**(983.00)**	**0.037***
Instructor Quality	37	4.92	0.28	68	4.81	0.43	(1136.50)	0.179
Summative Exam Performance:								
All Questions (59 questions)	63	82.80	14.57	97	82.63	13.47	(1668.50)	0.698
Lower Bloom's (50 questions)	63	85.06	12.61	97	84.72	11.50	(1181.00)	0.634
Higher Bloom's (9 questions)	63	70.22	18.85	97	71.00	18.08	(39.00)	0.894

Table 3. Description Statistics for Terms 3 and 4. At the start of each semester, Term 4 had 64 students enrolled and Term 3 had 101 students enrolled. One student withdrew from Term 4 and four students withdrew from Term 3. N = total number of student responses for each survey question. For exam performance, n = number of students that completed both midterm-1 and midterm-1. SD = standard deviation. Summative exam performance represents the average percent correct in the class for the questions analyzed. Mann U Whitney tests were used to compare responses between groups. P values are shown for each category. Bold categories were statistically significantly difference between the groups. (*) = p is less than 0.05, (**) = p is less than 0.01, and (***) = p is less than 0.001.

defined student learning objectives," "clear expectations for assignments and other work," or "perceptions of difficulty." In our second comparison between Terms 1 and 2 (19 and 103 students, respectively), the class size difference was even greater (5.4-fold increase) compared to Terms 3 and 4 (1.58-fold increase). Consistent with the greater size difference, three of the classroom environment measures were rated significantly higher by students in the lower enrollment term: course quality, instructor quality, and oral communication (Table 4).

In both of these comparisons, the higher enrollment terms had significantly lower course quality ratings. Interestingly, this was despite having the same course structure and content, as well as a similar proportion of class time devoted to interacting with peers and engaging in collaborative learning. The instructor was also the same, yet, in the Term 1 versus Term 2 comparison, the higher enrollment group had a significantly lower rating for instructor quality. These findings are in agreement with past studies describing a negative relationship between class size and course evaluation scores (Bedard and Kuhn, 2008; Benton and Cashin, 2012; Kwan, 1999; Mandel and Sussmuth, 2011; Monks and Schmidt, 2011; Sapelli and Illanes, 2016; Westerlund, 2008).

One possible explanation for the decreased student ratings of course and instructor quality with the higher enrollment TBL classes may be decreased student-faculty interaction and one-on-one attention. While we have not

directly measured the amount of student-faculty interaction in our course terms, past studies have documented fewer interactions and difficulty engaging with students in larger classes (Carbone and Greenberg, 1998; Mulryan-Kyne, 2010; Wulff et al., 1987). However, despite the additional challenges faced with high enrollment TBL classes, several measures of classroom environment did not change across different class sizes in our study. We propose that the frequent interactions with team members throughout the semester may have helped maintain the ratings for a 'welcoming and inclusive classroom environment.' However, given that our study did not include a non-TBL control group with the same instructor, we cannot be sure whether this rating of the classroom environment is a direct result of the TBL approach or perhaps, instructor personality and rapport with students.

Class Size and Exam Performance in TBL
Our finding that exam scores were similar across a range of enrollment sizes (Tables 3, 4, and 5) suggests that, for these TBL course terms, student exam performance (and presumably content knowledge outcomes) was not negatively impacted in a larger classroom environment. Furthermore, students' accuracy on questions that tested lower-order and higher-order cognitive outcomes did not significantly change with the size of the class (Tables 3, 4, and 5). The design elements of TBL that promote motivation and learning- retrieval practice, spaced practice,

Table 4. Student Ratings and Exam Scores for Terms 1 and 2

	19 Students (Term 1)			103 Students (Term 2)				
	n	mean	SD	n	mean	SD	(U)	p value
Classroom Environment:								
Learning Objectives	13	4.31	1.49	59	4.75	0.28	(347.50)	0.416
Clear Expectations	13	4.31	1.49	59	4.58	0.95	(467.50)	0.890
Oral Communication	**12**	**4.25**	**1.14**	**50**	**3.58**	**1.01**	**(174.50)**	**0.019***
Welcoming/Inclusive	13	4.31	1.49	59	4.44	0.91	(341.00)	0.056
Course Quality	**12**	**5.00**	**0.00**	**58**	**4.14**	**0.74**	**(120.00)**	**0.000***
Instructor Quality	**12**	**5.00**	**0.00**	**58**	**4.62**	**0.57**	**(228.00)**	**0.017***
Summative Exam Performance:								
All Questions (44 questions)	16	77.80	12.60	101	80.09	12.72	(813.00)	0.195
Lower Bloom's (37 questions)	16	77.59	12.82	101	81.00	12.13	(538.00)	0.113
Higher Bloom's (7 questions)	16	78.86	12.19	101	75.29	15.65	(21.50)	0.701

Table 4. Descriptive Statistics for Terms 1 and 2. Term 1 had 19 students enrolled and Term 2 had 103 students enrolled at the start of the term. Three students withdrew from Term 1 and two students withdrew from Term 2. n = total number of student responses for each survey question or total number of students completing both the first and second midterm exam. SD = standard deviation. Summative exam performance represents the average percent correct in the class for the questions analyzed. Mann U Whitney tests were used to compare responses between groups. P values are show for each category. Bold categories were statistically significantly different between the groups. (*) = p is less than 0.05, (**) = p is less than 0.01, and (***) = p is less than 0.001.

and feedback (Butler et al., 2014; Dunlosky et al., 2013; Schmidt et al., 2019)- would be present to the same extent in any-sized TBL class, thus providing a possible explanation for why exam performance did not vary across our different-sized TBL course terms.

Limitations and Future Directions
One limitation of this study was the need to include a summer session course in the analysis between Terms 1 and 2. Fall terms for this course typically enroll 60 or more students. As such, the summer session term provided the only opportunity to examine student performance and the classroom environment in a low enrollment version of this course. Additionally, the summer session version of this course was only offered one year, so additional replicates were not possible. As a result, we cannot rule out the possibility that the duration of the course terms- 6 weeks for summer session versus 14 weeks for fall/spring term courses- had an effect on the content knowledge outcomes. Indeed, while some studies show no effect, others have reported that shorter course terms are associated with improved performance on exams (Austin and Gustafson, 2006; Scott and Conrad, 1992). How this may manifest in TBL classes should be further explored.

The summer session TBL term enrolled a number of visiting and graduate/post-baccalaureate students. By comparison, the high enrollment version of the class offered during the fall term enrolled a low percentage of these non-degree seeking students (Table 2). To account

for the possibility that these differences in student demographics influenced our outcomes, we re-ran the analysis excluding the data from non-degree students (Table 5). Our findings were the same: there was no significant difference in exam performance between low and high enrollment TBL groups.

In order to generalize our findings, future studies should explore the relationship between TBL course size and student performance with a larger number of classes and broader range of courses at different institutions, as well as with different student populations. For example, it would be particularly valuable to know the content knowledge outcomes and students' perceptions of quality of the course and their instructor in introductory STEM classes which are frequently large and can have enrollments of 300 to 400. Monitoring classroom behaviors and faculty interactions to confirm that they are indeed reduced in larger TBL courses should also be further explored.

Proposed Strategies for Larger TBL Classes
Despite any drop in student exam performance, our higher enrollment course terms did suffer from lower evaluations of course quality. While this may be an unavoidable consequence of higher enrollment classes, there may be strategies that TBL facilitators can implement in large classes to improve student perception of the course and instructor. As is always the case with TBL, facilitators should work hard to achieve student buy-in with the TBL approach and should ensure that students learn how to

Table 5. Exam Scores for Terms 1 and 2, Excluding Visiting and Graduate Students

| | 19 Students (Term 1) | | | 103 Students (Term 2) | | | | |
	n	mean	SD	n	mean	SD	(U)	p value
Summative Exam Performance:								
All Questions (44 questions)	8	84.16	13.72	100	80.23	12.72	(807.50)	0.179
Lower Bloom's (37 questions)	8	82.89	12.90	100	81.16	12.04	(578.50)	0.250
Higher Bloom's (7 questions)	8	80.86	15.40	100	75.29	15.98	(18.00)	0.404

Table 5. Descriptive Statistics for Terms 1 and 2, excluding visiting and graduate students. Term 1 had 19 students enrolled and Term 2 had 1-3 students enrolled at the start of the term. n = total number of student responses for each question or total number of students completing both the first and second midterm exam, excluding visiting and graduate/post-baccalaureate students. SD = standard deviation. Summative exam performance represents the average percent correct in the class for the questions analyzed. Mann U Whitney tests were used to compare responses between groups. P values are shown for each category. Bold categories were statistically significantly different between the groups. (*) = p is less than 0.05, (**) = p is less than 0.01, and (***) = p is less than 0.001.

work and communicate in teams effectively (Thompson et al., 2007a; Thompson et al., 2007b). In fact, an entire TBL module could be dedicated to helping students learn more about TBL and why it is being used. It is also possible that team teaching or utilizing teaching assistants as additional facilitators could help increase student engagement and one-on-one attention. Lastly, as has been shown for large non-TBL courses, the instructor can help make classes feel smaller by learning names (Holstead, 2019), walking through aisles, using humor, and having a more engaging classroom presence (Cash et al., 2017). Future studies should investigate the impact of these instructor behaviors on student perception and enjoyment of TBL in a large classroom setting.

REFERENCES

Ake-Little E, von der Embse N, Dawson D (2020) Does Class Size Matter in the University Setting? Educational Researcher 49:595-605.

Allen RE, Copeland J, Franks AS, Karimi R, McCollum M, Riese DJ II, Lin AY (2013) Team-based learning in US colleges and schools of pharmacy. Am J Pharm Educ 77:115.

Arias J, Walker D (2004) Additional evidence on the relationship between class size and student performance. The Journal of Economic Education 35:311-329.

Austin A, Gustafson L (2006) Impact of Course Length on Student Learning. Journal of Economics and Finance Education 5:26-37.

Baeten M, Kyndt E, and KDochy, F (2010) Using student-centered learning environments to simulate deep approaches to learning: factors encouraging or discouraging their effectiveness. Educ Res Rev 5:243-260.

Ballen C, Aguillon S, Brunelli R, Drake A, Wassenberg D, Weiss S, Zamudio K, and Cotner S (2018) Do Small Classes in Higher Education Reduce Performance Gaps in STEM? BioScience 68:593-600.

Bandiera O, Larcinese V, Rasul I (2010) Heterogeneous class size effects: New evidence from a panel of university students. Economic Journal 120:1365-1398.

Becker WE, Powers JR (2001) Student performance, attrition, and class size given missing student data. Econ Educ Rev 20:377-388.

Bedard K, Kuhn P (2008) Where class size really matters: Class size and student ratings of instructor effectiveness. Economics of Education Review 27:253-265.

Bellante DM (1972) A summary report on student performance in mass lecture classes of economics. Journal of Economic Education 4:53-54.

Benton,SL, Cashin WE (2012) Student ratings of teaching: A summary of research and literature. IDEA paper. Available at https://www.ideaedu.org/Portals/0/Uploads/Documents/IDEA%20Papers/IDEA%20Papers/PaperIDEA_50.pdf.

Benton, S.L, and Pallett, W.H. (2013) Class size matters: Essay on importance of class size in higher education. Inside Higher Ed. Available at
 https://www.insidehighered.com/views/2013/01/29/essay-importance-class-size-higher-education.

Bloom B (1956) Taxonomy of educational objectives: The classification of educational goals. New York, NY: Longmans Green.

Butler A, Marsh E, Slavinsky J, Baraniuk R (2014) Integrating cognitive science and technology improves learning in a STEM classroom. Educational Research Review 26:331-340.

Carbone E, Greenberg J (1998) Teaching large classes: unpacking the problem and responding creatively. To Improve the Academy 17:311-326.

Cash CB, Letargo J, Graether SP, Jacobs SR (2017) An Analysis of the Perceptions and Resources of Large University Classes. CBE Life Sci Educ 16:1-12.

De Paola M, Ponzo M, Vincenzo S (2013) Class size effects on student achievement: Heterogeneity across abilities and fields. Education Economics 21:135-153.

Dunlosky J, Rawson KA, Marsh EJ, Nathan MJ, Willingham DT (2013) Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. Psychol Sci Public Interest 14:4-58.

Eddy SL, Hogan KA (2014) Getting under the hood: how and for whom does increasing course structure work? CBE Life Sci Educ 13:453-468.

Edgell J (1981) Effects of class size upon aptitude and attitude of pre-algebra undergraduate students. Education Resources Information Center Document 203760. Washington, DC: Institute of Education Sciences and U.S. Department of Education. Available at https://files.eric.ed.gov/fulltext/ED203760.pdf.

Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP (2014) Active learning increases student performance in science, engineering, and mathematics. Proc Natl Acad Sci U S A 111:8410-8415.

Gibbs G, Lucas L, Simonite V (1996) Class size and student performance: 1984-94. Stud High Educ 21:261-273.

Gleason J (2012) Using technology-assisted instruction and assessment to reduce the effect of class size on student outcomes in undergraduate mathematics courses. College Teaching 60:87-94.

Haidet P, Kubitz K, McCormack WT (2014) Analysis of the Team-Based Learning Literature: TBL Comes of Age. J Excell Coll Teach 25:303-333.

Hancock TM (1996) Effects of class size on college student achievement. College Student Journal 30:479-481.

Hill MC (1998) Class size and student performance in introductory accounting courses, further evidence. Issues in Accounting Education 13:47-64.

Holstead C (2019) Want to improve your teaching? Start with the basics: learn students' names. The Chronicle of Higher Education. Available at https://www.chronicle.com/article/want-to-improve-your-teaching-start-with-the-basics-learn-students-names/.

Kennedy PE, Siegfried JJ (1997) Class size and achievement in introductory economics: evidence from the TUCE III data. Econ Educ Rev 16:385-394.

Kogl Camfield E, McFall EE, Kirkwood ML (2016) Leveraging innovation in science education: Using writing and assessment to decode the class size conundrum. Liberal Education 101/102. Available at https://www.aacu.org/liberaleducation/2015-2016/fall-winter/camfield.

Kokkelenberg EC, Dillon M, Christy SM (2006) The effects of class size on student grades at a public university. Cornell Higher Education Research Institute. Available at http://digitalcommons.ilr.cornell.edu/workingpapers/66/.

Kwan K (1999) How fair are student ratings in assessing the teaching performance of university teachers? Assessment and Evaluation in Higher Education 24:181-195.

Mandel P, Sussmuth B (2011) Size matters. The relevance and Hicksian surplus of preferred college class size. Econ Educ Rev 30:1073-1084.

Matta BN, Guzman JM, Stockly SK, Widner B (2015) Class size effects on student performance in a Hispanic-serving institution. The Review of Black Political Economy 42:443-457.

Michaelsen L, Knight A, Fink L (2004) Team-based learning: A transformative use of small groups in college teaching. Sterling, VA: Stylis Publishing.

Michaelsen L, Sweet M (2008) The essential elements of team-based learning. New Dir Teach Learn 116:7-27.

Michaelsen L, Watson W, Cragin J, Fink L (1982) Team Learning: A Potential Solution to the Problems of Large Classes. The Organizational Behavior Teaching Journal 7:13-22.

Monks J, Schmidt RM (2011) The impact of class size on outcomes in higher education. The BE Journal of Economic Analysis & Policy 11:1-17.

Mulryan-Kyne C (2010) Teaching large classes at college and university level: challenges and opportunities. Teach High Educ 15:175-185.

Ng M, Newpher T (2020) Comparing Active Learning to Team-Based Learning in Undergraduate Neuroscience. Journal of Undergraduate Neuroscience Education 18:99-108.

Olson JC, Cooper S, Lougheed T (2011) Influences of teaching approaches and class size on undergraduate mathematical learning. PRIMUS 21:732-751.

Raimondo H, Esposito L, Gershenberg I (1990) Introductory class size and student performance in intermediate theory courses. The Journal of Economic Education 21:369-381.

Sapelli C, Illanes G (2016) Class size and teacher effects in higher education. Econ Educ Rev 52:19-28.

Schmidt, H.G, Rotgans, J.I, Rajalingam, P, and Low-Beer, N. (2019) A Psychological Foundation for Team-Based Learning: Knowledge Reconsolidation. Acad Med 94:1878-1883.

Scott P, Conrad C (1992) A Critique of Intensive Courses and an Agenda for Research. In Higher Education: Handbook of Theory and Research 8:411-459.

Thompson BM, Schneider VF, Haidet P, Levine RE, McMahon KK, Perkowski LC, Richards BF (2007a) Team-based learning at ten medical schools: two years later. Med Educ 41:250-257.

Thompson BM, Schneider VF, Haidet P, Perkowski LC, Richards BF (2007b) Factors influencing implementation of team-based learning in health sciences education. Acad Med 82:S53-56.

Westerlund J (2008) Class size and student evaluations in Sweden. Education Economics 16:19-28.

Wright MC, Bergom I, Bartholomex T (2019) Decreased class size, increased active learning? Inteded and enacted teaching strategies in smaller classes. Active Learning in Higher Education 20:51-62.

Wulff DH, Nyquist JD, Abbott RD (1987) Students' perceptions of large classes. New Dir Teach Learn 32:17-30.

Address correspondence **to:** Dr. Thomas Newpher, Department of Psychology and Neuroscience, 308 Research Drive, Duke University, Durham, NC 27708. Email: thomas.newpher@duke.edu

APPENDIX

Example questions from summative exams

Bloom's lower level - Recall
The separation of oppositely charged ionic particles across a resting neuron's membrane results in a potential that is measured as:
A. current.
B. voltage.
C. conductance.
D. resistance.
E. permeability.

Bloom's lower level – Understand
Which statement about the image to the left is true?
A. The sensory neuron indicated by the label C is demonstrating convergence and only resides within the peripheral nervous system.
B. The interneuron indicated by the label B is found in both the peripheral and central nervous systems.
C. The motor neuron indicated by the label A is demonstrating divergence and is found in both the peripheral and central nervous systems.
D. The sensory neuron indicated by the label A is demonstrating divergence and is found in both the peripheral and central nervous systems.
E. The interneuron neuron indicated by the label D resides in both the peripheral and central nervous systems.

Bloom's higher level – Apply
Hypokalemia is a serious medical issue characterized by low serum potassium levels. For treatment, patients are often given an IV containing K+ ions. Before treatment, the resting membrane potential in a general neuron in the brain is _____, and once treatment is finished and standard physiological conditions are restored, the RMP is -65 mV.
A. around -85 mV
B. around -45 mV
C. around 0 mV

Bloom's higher level - Analyze
From the data shown in this figure we could conclude that:
A. RTX is a toxin that only kills second order nociceptive neurons.
B. Shank3 is not present in the dorsal horn of the spinal cord.
C. Some amount of Shank3 is present in the presynaptic terminal of first order nociceptive neurons.
D. Shank3 is highly enriched in the cell bodies of DRG neurons, but not present in the presynaptic terminal.

Bloom's higher level - Evaluate
Which strategy would be the best method to treat Alzheimer's disease, if your goal is to limit the number of undesirable side effects?
    A. addition of a drug that prevents AB binding to PrPc
    B. addition of a drug that prevents endocytosis of AMPARs
    C. gamma secretase inhibitors
    D. Fyn kinase inhibitors
    E. NMDAR antagonists
    F. protein translation inhibitors