Examining the Effectiveness of Team-Based Learning (TBL) in Different Classroom Settings

Richard F. Yuretich[1,a] and Lisa C. Kanner[1]

ABSTRACT

The problem of effective learning in college classrooms, especially in a large lecture setting, has been a topic of discussion for a considerable span of time. Most efforts to improve learning incorporate various forms of student-active learning, such as in-class investigations or problems, group discussions, collaborative examinations and interactive technology such as i-clicker. Team-Based Learning (TBL) is a classroom paradigm that all but eliminates the standard lecture and shifts exams from an end-of-unit exercise to a preparatory role. Adapting TBL to an introductory oceanography course has yielded mixed results. Student satisfaction as measured by course evaluations has increased in several important indicators, but not in summative questions. Final grades have also continued at high levels, and discussions around in-class investigations have become increasingly sophisticated, but performance on certain exam components associated with higher-order learning has declined. The classroom environment may play a major role, with TBL showing greater effectiveness in a TBL-designed facility compared with a traditional auditorium. © *2015 National Association of Geoscience Teachers*. [DOI: 10.5408/13-109.1]

Key words: team-based learning, active learning, assessment, higher-order learning

INTRODUCTION

The past twenty years has seen a concerted effort to increase the use of active-learning methods in STEM courses at the college level. Among these techniques are collaborative or cooperative learning (Johnson et al., 1998), project-based learning (Regassa and Morrison-Shetlar, 2009) and problem-based learning (Allen, 1997). There is now ample evidence that these methods promote critical thinking, higher-order processing and greater retention of information than lecture-only instruction (Springer et al., 1999), yet a number of substantive issues about the effective implementation of active-learning methods remain (Prince, 2004). In order for active-learning methods to succeed, it is essential for students to come to class prepared; therefore, the development of mechanisms to accomplish this are essential.

Team-Based Learning (TBL) is a robust variety of collaborative learning that confronts this issue using a highly structured framework (Michaelsen et al., 2002). It was first implemented in classes for management and social sciences and it is now being used in many different subject areas, including the natural sciences (Metoyer et al., 2014). In this article, we discuss our experiences in teaching two large-enrollment (90 and 300 students) oceanography classes using TBL and we evaluate the effects that the method has upon student learning and achievement.

The TBL model

In order to prepare students to engage in active learning and collaborative investigative projects, Team-Based Learning is built around four main components. The first component is the formation of permanent teams. These are relatively large by the standards of "traditional" cooperative learning. Five to seven students per team is the recommended size (Michaelsen et al., 2002). The rationale for the large size is that the team can continue to function effectively as a unit even if a few members are absent. These teams are assembled by the instructor with the goal of balancing skills, experience, content knowledge, and diversity among the teams. Formal roles for each team member are not assigned so that the team may use each member's skills most effectively.

The second major component of TBL is the "Readiness Assurance Process." In the TBL paradigm, the delivery of information during class time is minimized or even eliminated, so that students must complete assigned readings on their own to be ready for investigations during class. The motivation for this preparation is the Readiness Assessment Test (RAT). These tests are administered at the start of a unit or module and comprise three phases in a manner similar to two-stage cooperative or pyramid exams (Yuretich et al., 2001). The first phase is an individual closed-book test that can be completed in a short period of time. This is followed by a repeat of the test that is completed as a team and it is also closed-book. The third phase of the RAT is the "challenge" that allows students to propose alternate correct answers based on evidence they can find in their readings or notes. The success of the RAT requires grading of the individual and team tests in real time so students will have feedback for the challenge portion.

Team investigations, the third component of the TBL format, are the centerpiece of the method. The secret to the success of the effort is that teamwork or collaboration is required only during class time to avoid the myriad problems in having students arrange meetings around their own often conflicting schedules. The in-class team investigations revolve around a topic that builds upon the information acquired during the Readiness Assurance Process. All teams work on the same investigation so they can share insights during the reporting phase. This team reporting process is also most effective if it can be done in a way that allows teams to present their answers simultaneously, so that the

Received 11 December 2013; revised 30 October 2014 and 23 February 2015; accepted 12 March 2015; published online 13 May 2015.
[1]Department of Geosciences, University of Massachusetts, 611 N Pleasant St., Amherst, Massachusetts 01003, USA
[a]Author to whom correspondence should be addressed. Electronic mail: yuretich@geo.umass.edu. Tel.: 413-545-0538

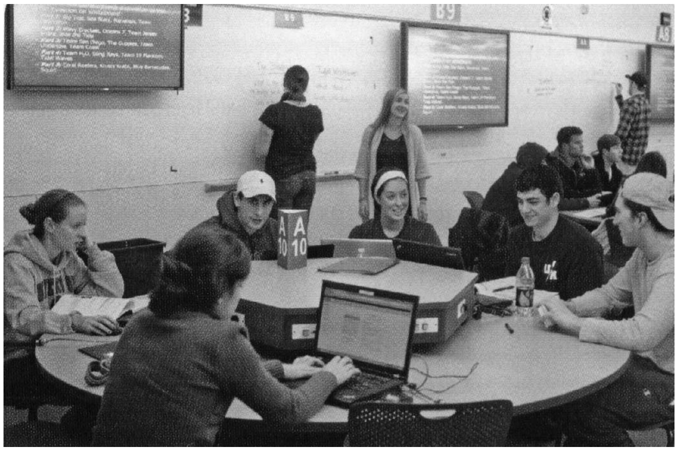


FIGURE 1: TBL-dedicated classroom showing students engaged in discussion and displaying answers to investigations on whiteboard.

teams can share information and ideas among one another. Individual accountability within the team effort requires an effective system of peer evaluation and this comprises the fourth major component of TBL.

METHODS OF IMPLEMENTATION

Research Tools

Comparison With Legacy Course

Team-Based Learning was adapted to oceanography at UMass-Amherst using the foundation for active learning established in this course for several years (Yuretich et al., 2001; Yuretich, 2003). This course enrolled 300 students in an auditorium setting. There are no laboratories or separate discussion sessions. The course was divided into five topical modules, each of which is concluded by a two-stage cooperative exam.

Class Size and Classroom Design

The initial implementation of Team-Based Learning (which we will call "TBL-1") occurred in a class of 90 students in a specially-designed TBL classroom. This classroom has round tables seating nine students apiece and each table has access to its own video monitor and whiteboard (Fig. 1). The second trial (TBL-2) was conducted in a traditional lecture hall seating 300 students in fixed, stadium-style seats. This was the same classroom that had been used for earlier versions of the course.

Modified Readiness Assurance Process

The prescribed Readiness Assurance Process was modified slightly from the established procedures. Readings were assigned prior to each instructional module and encompassed anywhere from 30 to 50 pages of text and figures. The first class of each module consisted of a "concept preview" to highlight the key elements of the reading. Each team was given a focus question to develop and then write down their collective answer (Fig. 2). The questions and answers were then presented for discussion among the entire class.

In-class Team Investigations

After the RAT was completed, each remaining class session in the module was devoted to completing and discussing a team investigation. The instructors provided a 10-minute overview or lead-in to the planned exercise. The investigation itself was generally done on paper, finished during the class period, with one submission per team (Fig. 3).

Peer Evaluation

Peer evaluation is an essential piece of a TBL course that provides an effective degree of individual accountability for team assignments (Cestone et al., 2008; Doyle and Meeker, 2008). We used an online evaluation program (iPeer®) that alleviated some of the stress associated with this task.

On-line Learning Management System

The "standard" TBL components were augmented by an on-line learning management system (Moodle®) to further encourage exploration and individual accountability. This was achieved through a moderately challenging quiz that had to be completed by the end of each module.

Student Course Evaluations

The numerical scores on selected questions of the standard end-of-semester course evaluations were com-

Concept Preview for Part 5

A. Waves

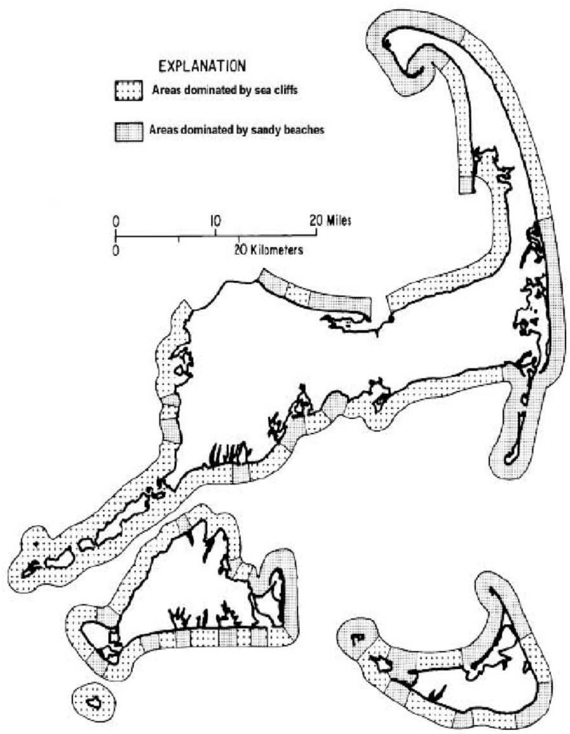
1. How do wind waves form?
2. What is wave length? Wave height? Wave period?
3. How does water move within deep-water waves?
4. What determines the celerity (velocity) of deep-water waves?
5. Under what conditions do whitecaps or breakers form?
6. What is an intermediate or shallow-water wave? What determines the celerity (velocity) of these waves?

FIGURE 2: Example of questions used for "concept preview" prior to a RAT.

Investigation 5.6 Coastal Erosion

During more than 350 years of European settlement, the coast of Cape Cod has been inundated by strong storms 84 times! During the winter of 1978, and again during the "Perfect Storm" of 1991 when powerful nor'easters pounded the coast. The effects were widespread and devastating, but out of sight and out of mind of many "summer people".

1). The map below shows the location of cliffs and sandy beaches along the Cape Cod shoreline. Identify the areas where erosion is the dominant process and explain why.



2). Sand dunes are a common feature on Cape Cod and other barrier beaches. Where are dunes usually located with respect to the beach? Can they be of value in preventing beach erosion?

FIGURE 3: Example of an in-class investigation completed as a team effort. Copyright © 2012 The McGraw-Hill Companies; reprinted with permission.

pared between the TBL and Legacy courses. In addition, the written comments were coded to judge the effectiveness of the changes in the course.

Student Performance

Grades can be influenced by a number of factors that are difficult to control, but in conjunction with other indicators they can be used as a measure of student learning. In this study, we compared the end-of-semester grades between the TBL and Legacy courses. We also did a closer analysis of the answers to specific questions on the final exams in order to gauge the level of higher-order thinking.

Analysis of Research Tools

Comparison With Legacy Course

The comparison here is made between the two offerings of the TBL course with a previous version of the course taught by the senior author in 2010 (called "2010 Legacy"). This course of 300 students in a traditional auditorium included many of the active-learning techniques that were incorporated into the TBL versions including in-class investigations, cooperative exams, and on-line homework. The in-class investigations in the Legacy 2010 course were done as "think-pair-share" or informal groups. Each student submitted his or her own work individually. The cooperative exams in the Legacy course were done as described in Yuretich et al. (2001) with a solo multiple-choice exam that was repeated as a collaborative effort. The cooperative portion of the final exam was made available to the students on-line after the conclusion of the classroom test. Online homework was also used to create more in-depth questions that probed more deeply into students' comprehension and problem-solving abilities. All these components were adapted for the TBL versions of the course.

Class Size and Classroom Design

The differences in the classroom design and size necessitated some accommodations in an effort to keep the teaching methods aligned between the two trials. Adaptations to the different classroom facilities are described in the appropriate subsections. Team selection is a critical component for TBL and similar methods were used for both courses to ensure heterogeneous grouping within teams (Metoyer et al., 2014). Students were initially stratified by class level, science confidence, and geography, then they counted off according to the number of teams for the class (20 in TBL-1, 52 for TBL-2).

Modified Readiness Assurance Process

The concept previews were used to stimulate discussion prior to the RATs. In the TBL-1 course, each team wrote the response to the "concept preview" question on the team whiteboard or video display. The projection system in the classroom then facilitated the global review of the question. For TBL-2 in the lecture hall, multiple teams were assigned the same question and representatives of the teams came to the front of the class to present their answers using the document camera.

The RATs took place during the next class period. Questions were multiple-choice and were derived from the major concepts discussed during the preview session. The RAT used optical scanning forms that were graded by a portable scanner in the classroom. The class was given 20 to 30 min for the individual portion of the RAT, then the answer sheets were collected and each team given one blank opscan sheet for the team RAT. While the team is discussing their collective answers, individual RATs are graded and returned to the teams at the conclusion of the team RAT. The challenge phase of the RAT often yielded lively discussions, especially in the larger section of the course which had a longer class period. TBL-1 was taught in 50-min sessions three times per week, whereas TBL-2 met twice weekly for 75 min. The latter schedule afforded more time for the RAT and led to more thorough discussions during the challenge component.

In-class Team Investigations

The team investigations replaced the activity-plus-short-lecture format used in the 2010 Legacy course. Depending on the nature of the investigation, the teams spent 10 to 20 min in discussing the problem among themselves. For TBL-1 the teams posted their results on either the whiteboards or their video screen which facilitated the communal discussion and review. TBL-2 used a reporting system similar to that employed in the "concept previews." Some representative teams (three to six) were chosen to go to the front of the room and display their responses using the document camera.

Each team handed in a single report for their investigation. The advantage of this is threefold. First, a single report encourages teams to work collaboratively. Second, the quality of the reports is generally better and more thoughtful than is typical of those generated individually. Third, the number of reports is reduced to a manageable level such that the answers could be read thoroughly and assessed quickly. In TBL-1, there were 20 reports to read; in TBL-2, this number increased to 52. In both cases it was relatively easy to read through them and determine what issues were grasped and what problems remained. In the manner of "Just-in-Time Teaching" (Novak, 2011), these issues were discussed further at the beginning of the next class meeting. Grading was done using a rubric designed for the purpose that also facilitated rapid assessment (Fig. 4).

Peer Evaluation

Peer evaluations can often be difficult and unreliable, but the iPeer® software provides a means to ease the process. First, students could do their team evaluation in the privacy of their own thoughts, away from the prying eyes of their erstwhile teammates. Secondly, the iPeer software gives each student a very useful tool to distribute the perceived effort of each teammate numerically (Fig. 5). In addition, brief written comments provided by the students proved to be very helpful in evaluating the individual and collective effort of the team. Results of the iPeer surveys can be verified by the instructor's own observations of the functioning of teams during class time.

On-line Learning Management System

The quizzes and homework assignments administered using the LMS can probe more deeply into higher-order thinking than high-stakes in-class tests. In general, the questions focused on calculations, data analyses, and interpretations of diagrams to go beyond the multiple-choice format of the RAT. Students were given two chances to do these quizzes, with some formative feedback on their initial response built-in to the quiz questions and answers to guide their second attempt. Calculation questions were programmed to change the values for the second attempt in order to eliminate the opportunity to "plug and chug" the answer from the first iteration. Grades were based on the better of the two trials. The final exam was also a traditional comprehensive multiple-choice test, but with a twist. When the official exam time was finished, the same test opened on Moodle and students had 48 hr to complete the second version if they chose (almost all did). Collaboration with their team was encouraged and it was open-book, but each student had to submit his or her own answers. Final exam grades were determined on some proportion of the in-class (60%) and online (40%) efforts.

Student Course Evaluations

The TBL-1 course used the standard evaluation form for UMass-Amherst (SRTI – Student Response to Instruction). The survey for TBL-2 was developed specifically for this class. We asked different questions than the standard course evaluation, since the latter focuses on the instructor and our goal was to probe the learning of the students more fully and to gauge the acceptance of the TBL method. This survey for TBL-2 was conducted on-line using the Moodle LMS and was not administered during class time. The completion rate of 97% is much higher than typical.

Student Performance

To examine the learning more closely, we compared the performance on the traditional (solo) part of the final exam among the classes. These multiple-choice examinations were all similar among the three course offerings and performance on this exam is a reasonable measure of knowledge base and ability to process information from the course. The students who enrolled in the course all came from the same

grade	criteria
10	logic of solution is clearly displayed; factual information is correct; units of measurement are correct and carried though the equation; calculations free of errors; answers are correct and have proper precision; conclusions are accurate
9	logic of solution may have minor lapses; units are usually correct; factual information is mostly accurate; calculations may have minor errors; answers or conclusions are essentially correct within a reasonable deviation
8	logic of solution may have uncertain components; units are mostly correct, although not always clear; some factual information is missing and documentation may be absent; calculations show noticeable errors; answers stray from desired values; conclusions deviate from desired outcome
7	logic of solution difficult to follow; units not always correct or shown; factual information is not correct; calculations have significant errors; answers not always within the realm of reasonable values; conclusions are not entirely justified
6	logic of solution is unclear; units are missing or incorrect; calculations have large errors; answers or conclusions are unreasonable.
5	Assignment done but large portions are incomplete.

Criteria Analysis

Logic ______________
Information ______________
Units ______________
Calculations ______________
Answers ______________
Conclusions ______________

Comments:

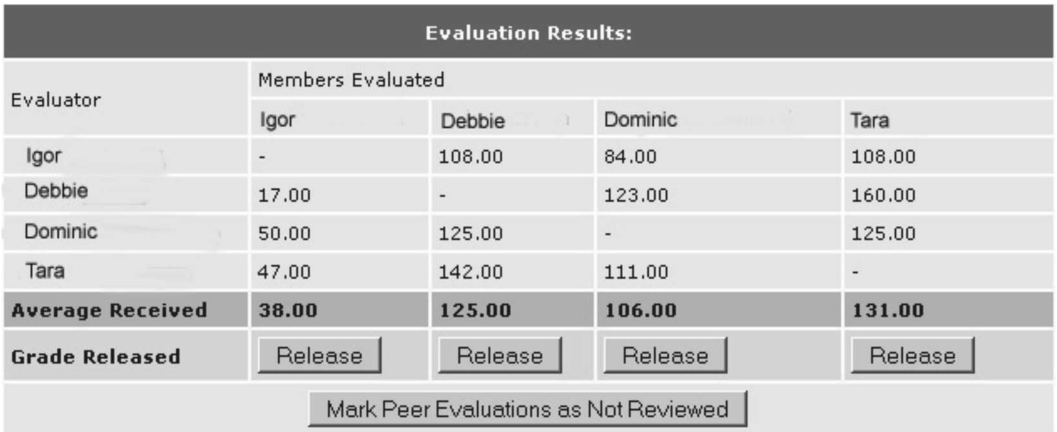
FIGURE 4: The scoring rubric used for rapid assessment of the team investigations.

population dominated by first-semester students, although average combined SAT scores for the student cohort in the course show slight increases for the three years (2010: 1167; 2011:1189; 2012:1197).

RESULTS

Student Responses

The reaction to TBL was generally favorable. The TBL-1 class in general showed a higher satisfaction level on all questions compared with the previous editions of the course (Fig. 6). A few of the most important qualities were significantly better at the 99% level. For example, the TBL-1 class overall agreed that their participation was stimulated much more than students in the 2010 legacy course (average score 4.3 out of 5 for TBL vs. 3.9 for legacy course). Even more of an increase came in the amount of useful feedback that the students felt they received (4.0 vs. 3.3). These are very important results given the goals of the TBL method. Free-form comments indicated that TBL was the most positive aspect of the course (Table I). Instructional technology elicited an equal number of positive and negative



Evaluation Results:				
Evaluator	Members Evaluated			
	Igor	Debbie	Dominic	Tara
Igor	-	108.00	84.00	108.00
Debbie	17.00	-	123.00	160.00
Dominic	50.00	125.00	-	125.00
Tara	47.00	142.00	111.00	-
Average Received	**38.00**	**125.00**	**106.00**	**131.00**
Grade Released	Release	Release	Release	Release
	Mark Peer Evaluations as Not Reviewed			

FIGURE 5: Example of the rating system in the iPeer® software used for team evaluations.

comments. The negative comments were mostly concerned about delays caused by equipment issues. This was the first time that the classroom had been used, so presumably this problem will resolve itself over time. On the other hand, the RATs elicited more negative than positive commentary. Most students viewed the RATs as a truly radical departure from the systemic norm that they had experienced in prior educational settings. A representative comment: "I thought RATs were a strange concept. It makes more sense to take a test after learning something about it in class."

The TBL-2 class in a standard lecture hall setting had a less positive view of the TBL experience as measured by the question concerning the overall rating for this course (Fig. 7). On the other hand, this rating is not significantly different from the averages of legacy versions of this course taught in the same room by the same instructor. All of the course components (readings, RATs, investigations, LMS) elicited favorable numerical rankings in learning effectiveness. Written comments were required as part of the TBL-2 survey, and Team-Based Learning elicited the greatest number of positive comments (123) (Table I). Some representative responses: "Team learning: it was much easier learning in teams with help of others than by myself." "I think that team based learning helped a lot to understand

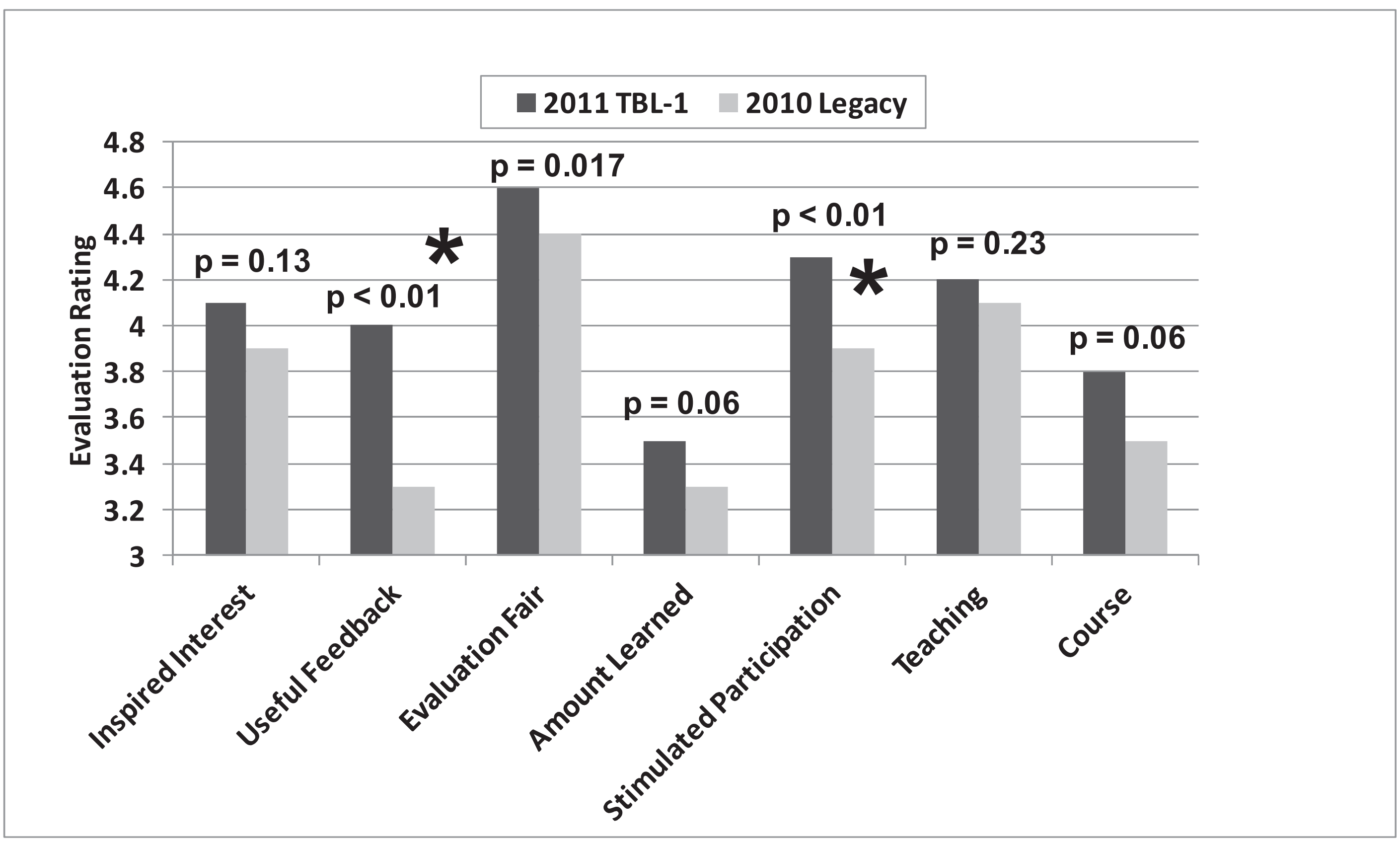


FIGURE 6: Comparisons of responses to questions on student course evaluation (UMass SRTI – Student Response to Instruction). Columns with asterisks represent significant differences at 99% level.

TABLE I: Items mentioned in free-form comments in course evaluations.

Course Component[1]	TBL-1 Positive	TBL-1 Negative	TBL-2 Positive	TBL-2 Negative
Team-Based Learning	24	1	123	30
Instructor	16	1	10	8
Technology	7	6	—	—
Investigations	2	2	9	6
Readiness Assessment Tests	1	7	51	15
Online component	1	1	12	9

[1]Other comments: RATs at end of unit; more lectures and note-taking; course content.

the material." "I enjoyed learning and double checking my answers with my team." "I did like that in your groups you always had someone to help clarify if you did not understand something." On the other side, there were 30 negative comments about teams or groups. Many of these came from students who were not thrilled by the collaborative nature of the course: "This course should just be a lecture course and not a TBL course." "I think you should lecture more in class. I felt that this was kind of a teach-yourself class and I didn't like that." Also the logistical difficulty of working with a relatively large team in a lecture hall with fixed seats was noted by several students: "The formation of the class made it hard to communicate with the other group members that were three rows ahead of me." "Unfortunately, the lecture hall wasn't helpful for a team setting. It was hard to hear others and participate correctly." In the comments, RATs were mentioned in a positive light 51 times, usually in connection with the collaborative aspect and the increased learning and grade improvement associated with them. In contrast there were 15 negative comments, most related to the administration of the RAT at the beginning of the module.

Student Achievement

The TBL-1 course showed a grade distribution that is highly skewed to the high end of the scale (Fig. 8). More than half the class received a final grade of either A or A−, whereas in the 2010 legacy course that used a variety of active-learning techniques (Yuretich et al., 2001) these letter grades constituted about one-third of the total. The increase reflects the importance of the team investigations and the online assignments in the grading scheme. The investigations were generally done well by the teams with a median score of 92. Likewise, the ability to complete the on-line assignments more than once improved the scores for this component, also with a median of 92. These scores together accounted for some 50% of the final grade and the numerical grades were not scaled in any manner. The smaller class size and technology-enhanced environment may have contributed to the enhanced grade distribution for this iteration. The comparison between the 2010 Legacy course and the

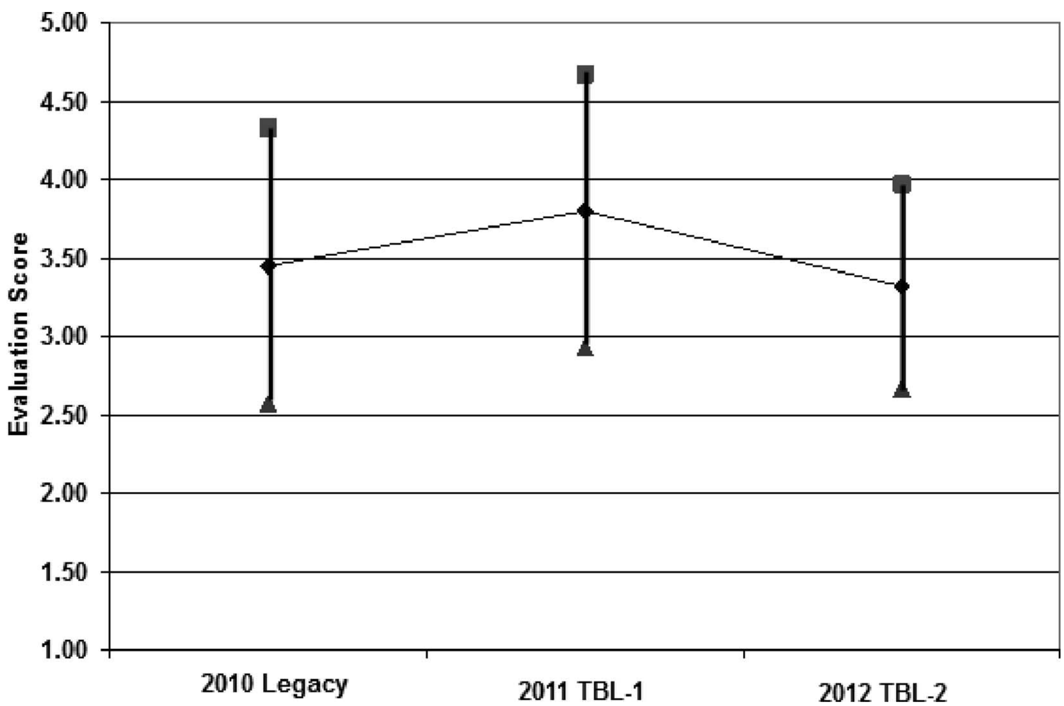


FIGURE 7: Comparisons of summative question (overall rating of course) from student evaluations from a representative legacy course (Fall, 2010) and the TBL-1 (Fall, 2011) and TBL-2 (Fall, 2012) oceanography. Line connects the average score; vertical bars represent one standard deviation.

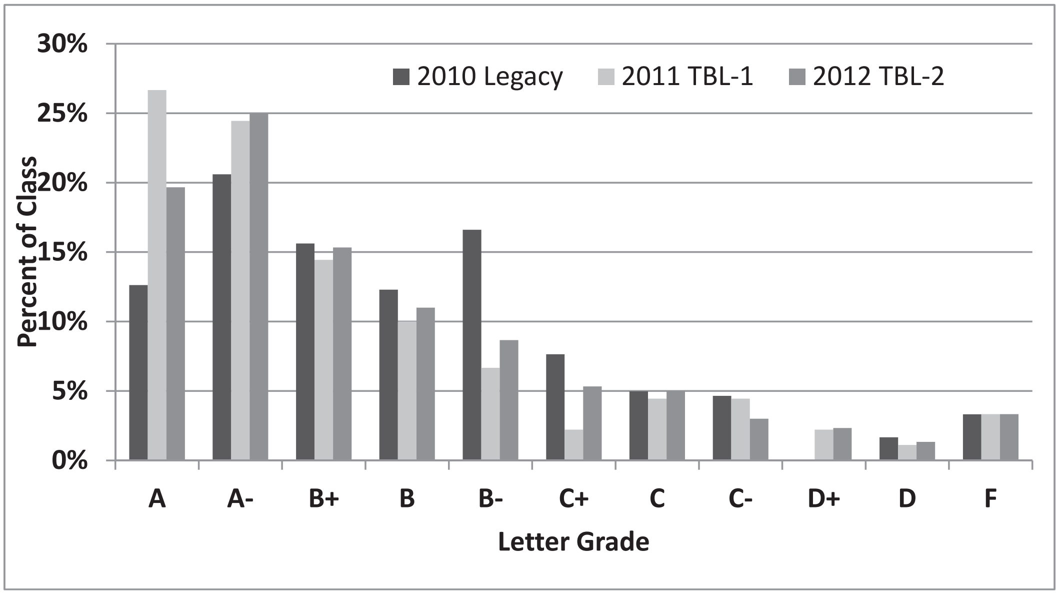


FIGURE 8: Comparison of final grade distributions among last three offerings of oceanography course by the senior author. "Legacy 2010" was a partial active-learning course with an enrollment of 301 students. "2011 TBL-1" is the team-based learning version in a technology-enhanced classroom with 90 students. "2012 TBL-2" is a team-based learning version in a traditional lecture hall with 300 students.

TBL-2 taught in the identical classroom with the same number of students also shows a noticeable improvement in the overall letter grades. There is a substantial and statistically significant increase in exam scores in the TBL-1 class, but the TBL-2 class had the lowest scores on the final exam in several years (Fig. 9). This relatively poor performance encompassed all components of the exam regardless of the nature of the topic or the sequence that it was discussed during the course. The classroom settings for the two TBL offerings may be a significant factor. Conducting the TBL-1 course in the dedicated, technology-enabled classroom created a very positive dynamic among the students and with the instructors. It was relatively easy for the students to communicate and report their findings, and all teams had the opportunity to report and respond. The physical layout of the traditional auditorium or lecture hall for TBL-2 restricted interaction among many teams, and the instructors were not able to listen in on most discussions. In addition, since the presentation of investigation results could only be done by a few teams in the large-class setting, the

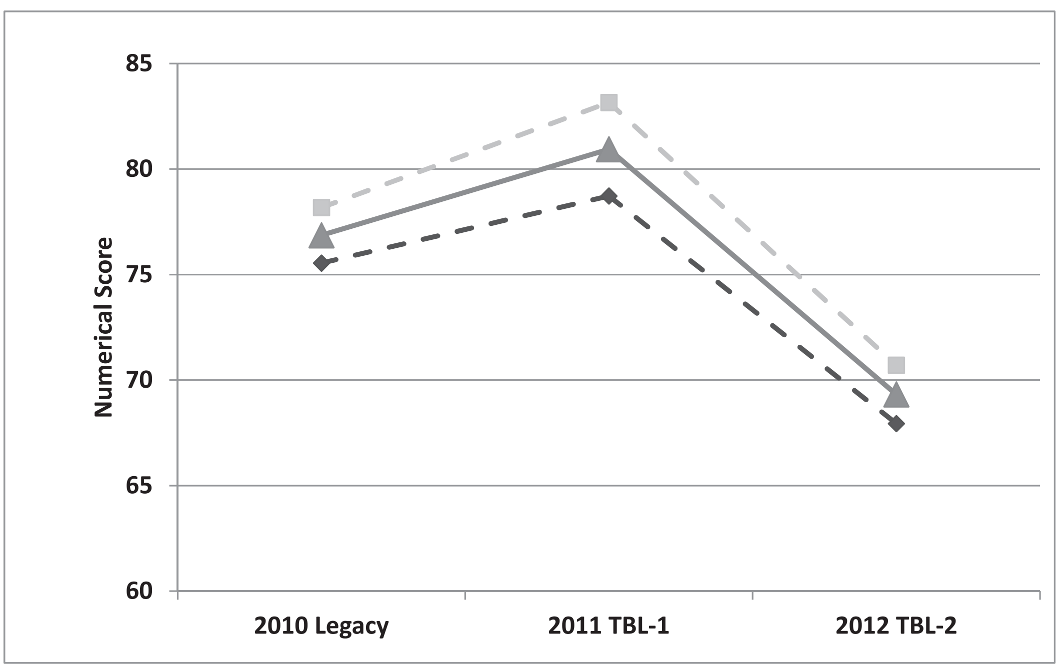


FIGURE 9: Grades on the solo portion of final exam in the three courses being examined. One standard deviation is shown around the mean and the differences are significant at the 99% level.

TABLE II: Comparison of student performance on final exam questions focused on higher-order reasoning (analysis, interpretation, synthesis). *t*-test assuming unequal variances.

	2010 (Legacy) vs. TBL-1		2010 (Legacy) vs. TBL-2		TBL-1 vs. TBL-2	
Mean	74.84615	63.92857	74.84615	53.23077	63.92857	53.23077
Variance	146.4744	477.6099	146.4744	555.5256	477.6099	555.5256
Observations	13	14	13	13	14	13
Hypothesized mean difference	0		0		0	
df	21		18		24	
t Stat	1.620628		2.941481		1.220335	
$P(T \leq t)$ one-tail	0.060009	(Null not	0.004363		0.117097	(Null not
t Critical one-tail	1.720743	rejected)	1.734064		1.710882	rejected)
$P(T \leq t)$ two-tail	0.120017		0.008726		0.234193	
t Critical two-tail	2.079614		2.100922		2.063899	

ability of the teams to completely process the results may have been diminished.

The goal of the team investigations is to promote higher-order reasoning skills and an important question is whether the TBL format has a measurable effect on this ability. Yuretich et al. (2001) found that students' solo performance on similar final exam questions that tested higher-order thinking (involving application, analysis, synthesis or evaluation according to Bloom et al., 1964) improved significantly when active-learning techniques were introduced into the lecture hall. The implementation of complete TBL does not show the same straightforward results (Table II). The mean scores for the higher-order reasoning questions are apparently lower in both TBL classes. However, the difference between the student performance in the legacy course (2010) and TBL-1 in the dedicated classroom is not statistically significant. The null hypothesis of no difference can also not be rejected at the 99% confidence level in the comparison between the two TBL sections (2011 and 2012), meaning that the students performed approximately the same. In any case, we cannot prove from these exam data alone that there has been a measurable increase in higher-order reasoning from the implementation of TBL.

DISCUSSION

There is now a deep literature confirming that active-learning methods lead to significant improvement in student comprehension (Mostrom and Blumberg, 2012). Team-Based Learning is also a collaborative technique that shows largely positive results when compared to direct-lecture classrooms (Carmichael, 2009). It is clear from our earlier studies that student achievement has been significantly better in this introductory oceanography course since active learning became an integral component. Student satisfaction is comparable between the earlier learner-centered course and the TBL versions as gleaned from course evaluations, and there are significant gains on some educationally beneficial aspects, such as feedback on assignments and fairness of grading. The organization of the class into permanent teams, arguably the most daunting and uncertain aspect from the instructor's viewpoint, was received positively by most students. The student performance level is generally comparable between the earlier active-learning classes and the TBL versions and grades skew toward higher ranges, yet the direct comparisons between final exam questions show that comprehension of some concepts does not seem to be helped by TBL, especially in the setting of a traditional auditorium-style classroom.

There is a long-standing debate concerning the loss of "content coverage" that occurs when active learning is implemented in a course, since the time devoted to instructor lectures is reduced. The standard TBL model eliminates practically all lecturing in favor of team collaboration during class time and the students get to explore aspects of the subject matter in a depth that may not be possible in a traditional classroom setting. Perhaps the absence of the instructor lectures puts students at some disadvantage for "connecting the dots" between the various team efforts. In this oceanography course, we tried to provide the requisite scaffolding by having the concept preview before the RAT to guide the students' reading and give an overview of the content in the module. One possible modification is to extend this concept preview into a second class period to allow more time for the students to process and reflect upon the information. Most students have been well trained to gain knowledge by listening and taking notes and a (small) subset are very good at this task.

The classroom environment may also have had an effect on the ability of all team members to participate in topical discussion. Cotner et al. (2013) found that students in active-learning classrooms outperformed expectations, in contrast to those who received the same instruction in traditional settings. Our results also indicate that TBL is more effective in active-learning classrooms with good communication systems.

After significant improvements in grades and test performance as a result of the implementation of active learning methods in oceanography (Yuretich et al., 2001, 2003), it is unclear whether further increases in achievement can be anticipated from a similar population of students There will presumably be a leveling off of performance indicators when consistent high achievement is reached that demonstrate the limits instructional methods can have upon learning.

The kinds of learning that are emphasized in TBL may not be documented effectively by traditional tests because of the somewhat reduced breadth of knowledge and because students in a TBL course are spending their time analyzing information and drawing conclusions. These efforts are best evaluated by the content of the completed in-class investigations. Using this as a measure, learning in a TBL setting is more substantial. Student answers to the questions posed by the investigations are more complete and the discussion surrounding them is more sophisticated. Such learning is not easily translatable into scores on multiple-choice tests. According to Straumanis (2012) successful learning can be characterized by three criteria: (1) long-term retention; (2) effective preparation for further or deeper learning and application; and (3) effective transfer of knowledge or skills to novel situations. Performance on final exams does not measure these aspects. The team-based investigations provide some insight but these are done collaboratively and cannot be used to assess individual knowledge or achievement. Developing an effective metric will be a research project in itself, perhaps by comparing the thoroughness and accuracy of the answers to the in-class investigations in a more traditional class to a TBL version to see if the differences in the two classes can be quantified more completely. In the meantime, our analysis confirms that learning-centered teaching methods, such as TBL, promote long-term retention of important ideas and concepts even though students in such classes do not always perform as well on tests of knowledge (Mostrom and Blumberg, 2012). Kim et al. (2013) investigated more completely the effects of active learning on students' critical thinking skills and concluded that there is a measureable improvement as a result, even though the level of their critical thinking remained at a ''developing'' stage.

CONCLUSIONS

Our experience indicates that TBL can be an effective framework for teaching science, especially in a setting that takes advantage of the opportunities for team investigation, discussion, and presentation. Dividing the class into permanent teams made the management and evaluation of student engagement more achievable. Average attendance during class sessions increased measurably to 80–84% in the TBL versions. The opportunity for the instructors to read and assess the student investigations after each class provides an excellent framework for addressing misconceptions or unresolved issues during the subsequent class, an important component of ''Just-in-time teaching'' (Novak, 2011). Student satisfaction improves noticeably in some components, especially in areas concerning feedback and assessment methods. Nevertheless, the effectiveness of TBL in promoting higher-order learning is uncertain and documenting this will require a more focused and controlled experiment. In general, there is a more positive outcome for TBL when it is conducted in a classroom that is dedicated to collaboration and has good technology support.

Acknowledgments

We are grateful to the Center for Teaching, Learning, and Faculty Development at the University of Massachusetts Amherst for financial support and encouragement through the TBL Faculty Fellows program. Karin Moyano Camihort was especially helpful in managing the program and providing substantive feedback on the course redesign and the original manuscript. The comments of two JGE reviewers and the associate editor helped in refocusing the content of the manuscript. This paper was completed while the senior author was serving at the National Science Foundation. All findings and opinions are the responsibility of the authors and do not reflect the views of the National Science Foundation.

REFERENCES

Allen, D.E., 1997. Bringing problem-based learning to the introductory biology classroom. *In* McNeal, and D'Avanzo, C., eds., Student-active science: Models of innovation in science teaching. Fort Worth, TX: Saunders, p. 259–278.

Bloom, B.S. 1964. Taxonomy of educational objectives: The classification of educational goals, by a committee of college and university examiners. New York: Longmans and Green.

Carmichael, J. 2009. Team-based learning enhances performance in introductory biology. *Journal of College Science Teaching* 38(4):54–61.

Cestone, C.M., Levine, R.E., and Lane, D.R. 2008. Peer assessment and evaluation in team-based learning. *New Directions in Teaching and Learning* 116:69–78.

Cotner, S., Loper, J., Walker, J.D., and Brooks, D.C. 2013. ''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.

Doyle, J.K., and Meeker, R.D. 2008. Team projects and peer evaluations. *College Teaching Methods & Styles Journal* 4(5):21–28.

Johnson, D., Johnson, R., and Smith, K. 1998. Cooperative learning returns to college: what evidence is there that it works? *Change* 30(4):26–35.

Kim, K., Sharma, P., Land, S., and Furlong, K. (2013). Effects of active learning on enhancing student critical thinking in an undergraduate general science course. *Innovative Higher Education* 38(4):223–235.

Metoyer, S.K., Miller, S.T., Mount, J., and Westmoreland, S.L. 2014. Examples from the trenches: improving student learning in the sciences using team-based learning. *Journal of College Science Teaching* 43(5):40–47.

Michaelsen. L.K., Knight, A.B., and Fink, L.D. 2002. Team-Based Learning: A transformative use of small groups in college teaching. Sterling, VA: Stylis Publishing. p. 286.

Mostrom, A.M., and Blumberg, P. 2012. Does learning-centered teaching promote grade improvement? *Innovative Higher Education* 37(5):397–405.

Novak, G.M. 2011. Just-in-Time Teaching. *New Directions for Teaching and Learning* 128:63–73.

Prince, M. 2004. Does active learning work? A review of the research. *Journal of Engineering Education* 93:223–231.

Regassa, L.B., and Morrison-Shetlar, A.I. 2009. Student learning in a project-based molecular biology course. *Journal of College Science Teaching* 38(6):58–67.

Springer, L., Stanne, M.E., and Donovan, S.S. 1999. Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: a meta-analysis. *Review of Educational Research* 69:21–51.

Straumanis, J. 2012. What we're learning about learning (and what we need to forget). *Planning for Higher Education* 40(4):6–12.

Yuretich, R.F. 2003. Encouraging critical thinking: measuring skills in large introductory science classes. *Journal of College Science Teaching* 33(3):40–45.

Yuretich, R.F., Khan, S.A., Leckie, R.M., and Clement, J.J. 2001. Active-learning methods improve student performance and scientific interest in a large oceanography course. *Journal of Geoscience Education* 49:111–119.