

The Impact of Team-Based Learning on Medical Students' Academic Performance

Paul G. Koles, MD, Adrienne Stolfi, MSPH, Nicole J. Borges, PhD, Stuart Nelson, PhD, and Dean X. Parmelee, MD

Abstract

Purpose

Since team-based learning (TBL) was introduced as a medical education strategy in 2001, few studies have explored its impact on learning outcomes, particularly as measured by performance on examinations. Educators considering implementing TBL need evidence of its effectiveness. This study was conducted to determine whether student performance on examinations is affected by participation in TBL and whether TBL benefits lower- or higher-performing students.

Method

The authors analyzed the performance of second-year medical students on 28

comprehensive course examinations over two consecutive academic years (2003–2004, 2004–2005) at the Boonshoft School of Medicine.

Results

The 178 students (86 men, 92 women) included in the study achieved 5.9% (standard deviation [SD] 5.5) higher mean scores on examination questions that assessed their knowledge of pathology-based content learned using the TBL strategy compared with questions assessing pathology-based content learned via other methods ($P < .001$, t test). Students whose overall academic performance placed them in the lowest quartile of the class benefited

more from TBL than did those in the highest quartile. Lowest-quartile students' mean scores were 7.9% (SD 6.0) higher on examination questions related to TBL modules than examination questions not related to TBL modules, whereas highest-quartile students' mean scores were 3.8% (SD 5.4) higher ($P = .001$, two-way analysis of variance).

Conclusions

Medical students' higher performance on examination questions related to course content learned through TBL suggests that TBL enhances mastery of course content. Students in the lowest academic quartile may benefit more than highest-quartile students from the TBL strategy.

Preclinical medical education is increasingly impaled on the horns of a curricular dilemma. While the volume of biomedical knowledge increases relentlessly, faculty–student contact

Dr. Koles is chair of pathology and associate professor, Departments of Pathology and Surgery, Boonshoft School of Medicine, Wright State University, Dayton, Ohio.

Ms. Stolfi is biostatistician and assistant professor, Department of Pediatrics, Boonshoft School of Medicine, Wright State University, Dayton, Ohio.

Dr. Borges is assistant dean for medical education research and evaluation, Office of Academic Affairs, and associate professor, Department of Community Health, Boonshoft School of Medicine, Wright State University, Dayton, Ohio.

Dr. Nelson was associate professor, Department of Pathology, Boonshoft School of Medicine, Wright State University, Dayton, Ohio; he is now retired.

Dr. Parmelee is associate dean for academic affairs and professor, Departments of Psychiatry and Pediatrics, Boonshoft School of Medicine, Wright State University, Dayton, Ohio.

Correspondence should be addressed to Dr. Koles, Boonshoft School of Medicine, Wright State University, 140 White Hall, 3640 Colonel Glenn Highway, Dayton, OH 45435-0001; telephone: (937) 775-2625; fax: (937) 775-2633; e-mail: paul.koles@wright.edu.

Acad Med. 2010;85:1739–1745.

First published online September 28, 2010
doi: 10.1097/ACM.0b013e3181f52bed

hours cannot be expanded in parallel. To address this dilemma, educators rely increasingly on textbooks, syllabi, electronic resources, and Web-based units of study that organize essential knowledge into accessible formats for independent learning outside class.^{1,2} As students acclimate to these learning tools, faculty may reduce hours previously reserved for lecture presentations, making face-to-face time more available for active teaching and learning strategies that engage learners and faculty in thoughtful dialogue and focus on application rather than acquisition of knowledge.

Problem-based learning (PBL), one such active learning method, has demonstrated its usefulness in undergraduate medical education during the past 40 years.³ PBL seems to be adaptable to changing curricular priorities. It was introduced as the primary learning strategy at McMaster University in 1969 and has survived two major revisions of the curriculum.⁴ In terms of outcomes, researchers have shown that medical students enrolled in a PBL curriculum demonstrate academic performance similar to students in a traditional lecture-based curriculum, as

measured by scores on United States Medical Licensing Examinations during a seven-year period.⁵

Team-based learning (TBL), which was introduced at Baylor College of Medicine in 2001, has a much shorter track record than PBL in medical education.^{6,7} Designed as an active learning strategy, TBL is learner-centered but instructor-led. It fosters individual and group accountability as small groups of students work together to answer questions.⁸ TBL employs a structured three-phase sequence: (1) preparation, during which learners study an advance assignment defined by faculty, (2) readiness assurance, where learners demonstrate knowledge through individual and group readiness assurance tests (RATs), and (3) application, when learners apply course concepts to problem-solving exercises designed by faculty and analyzed by teams.⁹

TBL's strategic sequence, when repeated multiple times during a course or academic term, encourages conscientious individual preparation while developing teams into cohesive learning groups. Faculty motivate students to thoroughly study the advance assignment by writing questions that assess mastery of critical

concepts in that assignment. These questions comprise the individual readiness assurance test (IRAT) and group readiness assurance test (GRAT). Both tests contain identical multiple-choice questions which are answered first by individual students, then by teams of five to seven students working together. Throughout the application phase, teams again collaborate to answer multiple-choice questions. During the readiness assurance and application phases, all teams simultaneously reveal their choices to the entire class.

TBL provides frequent opportunities for peers to enhance learning, as teammates talk and listen to one another to arrive at consensus decisions. Faculty invite teams to explain and support their choices publicly, and facilitate as teams debate justification for the best decision. Ideally, application questions require students to engage in critical thinking, rather than to merely retrieve relevant knowledge. Well-crafted application questions motivate teams to “make a concrete decision based on analysis of a complex issue.”^{10(p41)} Faculty often observe considerable energy and engagement of students during intra- and interteam discussions. Still, beyond the prospect of lively debate, an important question remains for educators: How effectively does TBL promote medical students’ learning?

Since TBL was introduced into medical education, few studies have correlated use of the method with students’ performance on examinations, particularly objective examinations that rely on multiple-choice questions to measure learning. Nieder and colleagues¹¹ showed no change in mean course examination performance compared with performance in years before TBL was used, but the use of TBL resulted in fewer students failing a human structure course. Levine and coinvestigators¹² found that third-year students in a psychiatry clerkship after TBL was implemented demonstrated significantly higher performance than did earlier cohorts on a National Board of Medical Examiners psychiatry subject exam. More recently, Letassy and colleagues¹³ reported that pharmacy students in an endocrine module achieved higher course grades after TBL replaced a lecture-based curriculum. A prospective study in 2002–2003 at the Boonshoft School of Medicine¹⁴ showed

no significant differences in performance by cohorts of second-year students on comprehensive course examinations (CCEs), regardless of whether they experienced TBL or case-based group discussion as a primary active learning method. However, both faculty and students in that study noted that TBL’s emphasis on individual preparation and peer-to-peer teaching seemed to enhance learning. The decreased failure rate observed in Nieder and colleagues’¹¹ study suggests that academically weaker students may benefit from TBL. This observation motivated us to investigate the effects of TBL on students across the full range of academic ability. Accordingly, we decided to examine learning outcomes for entire classes of students at the Boonshoft School of Medicine and for subgroups of students at both ends of the academic performance spectrum.

Educators who are considering implementing TBL into their curricula need objective evidence from studies that examine the impact of TBL on the learning outcomes of medical students. Given the small amount of such literature, the mixed results of previous studies of academic outcomes, and our own accumulated experience with TBL, we were guided in this study by two questions: (1) Does participation in TBL affect students’ performance on course examinations? (2) Does TBL preferentially benefit academically lower- or higher-performing students? We formulated two hypotheses: (1) Students will perform better on multiple-choice course examination questions if those questions are conceptually related to an advance assignment for a TBL module or to a TBL application exercise in that course, and (2) students whose academic performance places them in the lowest quartile of the class will benefit more from the TBL experience than will students in the highest academic quartile, as shown by comparison of each quartile’s performance on examination questions.

This study’s design differs from previous studies of TBL’s effectiveness in health professions education in an important way: Instead of comparing performance between different groups of learners in consecutive course iterations, we analyzed students’ performance *within* an academic year. This research design

decreases the confounding variables introduced by differences in academic ability between cohorts of students, changes in the roster of faculty and their teaching effectiveness, adjustments in course content or instructional methods, and variations in content or difficulty of examination questions. In this two-year study, students in each year’s cohort experienced the same composite of teaching faculty, educational strategies, course content, and multiple-choice examination questions.

Method

We examined the performance of second-year medical students on 28 major examinations over two consecutive academic years (2003–2004 and 2004–2005) at the Boonshoft School of Medicine. This study was deemed exempt by Wright State University’s institutional review board.

Boonshoft’s second-year curriculum

Boonshoft’s systems-based second-year curriculum consisted of 10 courses (divided between two terms) emphasizing foundational knowledge of physiology, pathology, and pharmacology applicable to clinical medicine. The sequence and content of these courses remained essentially stable over the study period. Teaching methods included lecture, laboratory exercises, clinical case discussions, independent study modules, and TBL modules; lecture was the method faculty used most frequently, but all classes included TBL modules. In all courses, students were primarily assessed via CCEs composed of multiple-choice questions and accounting for 80% to 95% of the overall course grade. Individual and group performance scores in TBL modules, including peer evaluations, accounted for 5% to 15% of the overall course grade. Three courses (neuroscience, blood, and respiratory) used additional graded assessments, accounting for <10% of the final course grade. Table 1 summarizes the academic terms, courses, CCEs, and TBL modules included in the study.

Teams of five to seven students were formed by random sorting at the beginning of the academic year; students remained on the same teams throughout all 10 courses. For TBL modules, advance assignments included readings from textbooks or journal articles, as well as

Table 1

Courses, Comprehensive Course Examinations (CCEs), and Team-Based Learning (TBL) Modules for the Second-Year Medical School Curriculum at Boonshoft School of Medicine, 2003–2005*

| Term and course | No. of weeks | Year | No. of CCEs | TBL modules: Topic (total no.) |
|--|--------------|------|-------------|--|
| Term 1 | | | | |
| Pathobiology and Therapeutics [†] | 3 | 2003 | 2 | Inflammation, genetic disorders, immune disorders, autonomic pharmacology (4) |
| | | 2004 | 2 | Inflammation, immune disorders, autonomic pharmacology (3) |
| Neuroscience | 8 | 2003 | 3 | Vascular disorders, central nervous system neoplasms, neurologic localization parts 1 and 2, neurodegenerative disorders (5) |
| | | 2004 | 4 | Vascular disorders, central nervous system infections, neurologic localization, neurodegenerative disorders (4) |
| Blood | 2 | 2003 | 1 | Peripheral blood cell interpretation (1) |
| | | 2004 | 1 | Peripheral blood cell interpretation (1) |
| Musculoskeletal and Skin | 2 | 2003 | 1 | Skin cancer (1) |
| | | 2004 | 1 | Skin cancer (1) |
| Term 2 | | | | |
| Cardiovascular | 4 | 2004 | 1 | ECG interpretation, cardiac pathology (2) |
| | | 2005 | 2 | ECG interpretation, cardiac pathology, cardiovascular autonomic pharmacology (3) |
| Respiratory | 3 | 2004 | 1 | Acute respiratory distress syndrome (1) |
| | | 2005 | 1 | Asthma (1) |
| Renal | 3 | 2004 | 1 | Urinalysis (1) |
| | | 2005 | 1 | Urinalysis (1) |
| Endocrine | 2 | 2004 | 1 | Multiple endocrine neoplasia (1) |
| | | 2005 | 1 | Autoimmune polyendocrine syndrome (1) |
| Reproduction | 2 | 2004 | 1 | Breast neoplasia (1) |
| | | 2005 | 1 | Breast neoplasia (1) |
| Gastrointestinal | 2 | 2004 | 1 | Liver disease (1) |
| | | 2005 | 1 | Liver disease (1) |

* Students' performance was assessed in 10 courses per year. Over the two years of the study, a total of 28 CCEs were administered and 35 TBL modules were used. ECG indicates electrocardiogram.

[†] This course covered general pathology and principles of pharmacology.

independent study tools created by faculty. About 60% of TBL modules' advance assignments included review of lecture content. Each module's RAT and application exercise were created by a faculty content expert and edited by at least one other member of the faculty. (A representative application exercise has been published elsewhere.¹⁵) TBL sessions were usually two hours long (40 minutes for RATs, 80 minutes for the application exercise) and facilitated by two members of the faculty, one of whom had created the module.

Multiple-choice questions for CCEs were authored and edited by numerous faculty representing multiple disciplines, in an attempt to create examinations that

assessed an integrated understanding of physiology, pathology, pharmacology, and clinical decision making. Two of the authors of this article (P.K. and S.N.) wrote approximately 50% of the pathology-based CCE questions; the rest were written by other pathology faculty members and edited by P.K. or S.N. Questions that were used in TBL modules did not appear on CCEs.

Data analysis

Without knowledge of students' performance on individual questions, we retrospectively analyzed all CCE questions on the 28 examinations students took over the study period to determine which questions required knowledge of pathology course content

to answer the question correctly. We identified these CCE questions as pathology-based questions (PBQs). We limited our study to PBQs because selection could be guided by two authors' area of expertise. PBQs were further divided into two subgroups by the author (P.K.) who designed or edited many of the TBL modules; he remained blinded to students' performance. One subgroup, designated TBL-related PBQs (TRs), contained questions that assessed knowledge included in a TBL module's advance assignment or discussed during a TBL application exercise. The second group, designated TBL-unrelated PBQs (TUs), consisted of questions that were conceptually unrelated to any TBL module's content.

Table 2

Classification of Comprehensive Course Examination (CCE) Questions: Relationship to Course Content in Pathology and Team-Based Learning (TBL) Modules, Second-Year Curriculum, Boonshoft School of Medicine, 2003–2005*

| Period | No. of CCEs | No. of TBL modules | Total no. of CCE questions | PBQs: No. (% of total) | TRs: No. (% of PBQs) | TUs: No. (% of PBQs) |
|--------------------------------|-------------|--------------------|----------------------------|------------------------|----------------------|----------------------|
| Academic year 2003–2004 | | | | | | |
| 2003, term 1 | 7 | 11 | 670 | 204 (30.4) | 62 (30.4) | 142 (69.6) |
| 2004, term 2 | 6 | 7 | 620 | 141 (22.7) | 52 (36.9) | 89 (63.1) |
| Academic year 2004–2005 | | | | | | |
| 2004, term 1 | 8 | 9 | 737 | 190 (25.8) | 65 (34.2) | 125 (65.8) |
| 2005, term 2 | 7 | 8 | 640 | 170 (26.6) | 64 (37.6) | 106 (62.4) |
| Total | | | | | | |
| 2 years, 4 terms | 28 | 35 | 2,667 | 705 (26.4) | 243 (34.5) | 462 (65.5) |

* PBQ indicates pathology-based exam question; TR, TBL-related PBQ; TU, TBL-unrelated PBQ.

We recorded every student's answer for each PBQ as 1 (correct) or 0 (incorrect) in an Excel spreadsheet. Only students who achieved scores in all CCEs and TBL modules for the entire academic year were included in our data analysis. The discrimination index of each PBQ was obtained from CCE item analysis as a useful indicator of question quality, and we calculated mean discrimination indices for TRs and TUs. We identified difficulty values for each PBQ (proportion of students answering that question correctly). We determined mean difficulty values and reported these as mean scores for TRs and TUs.

We compared the performance of all students on TRs versus TUs for all

courses combined, as well as for term 1 and term 2 courses separately, with paired *t* tests. We retrospectively classified students into four academic quartiles within their respective classes, based on cumulative performance on all graded assessments for the entire academic year, which allowed us to conduct discrete analysis of performance by the highest and lowest quartiles. We compared the performance of highest versus lowest academic quartiles on TRs versus TUs with a two-way analysis of variance (ANOVA), with quartile as an independent factor and question type as a repeated-measures factor. Scores are presented as mean percentage correct (standard deviation [SD]). We

considered *P* values <.05 to be statistically significant.

Results

Retrospective analysis of all multiple-choice questions from 28 CCEs produced 705 PBQs. These 705 PBQs accounted for 26.4% of all CCE questions during our two-year study (Table 2). Further classification of those 705 PBQs yielded 243 TRs (34.5%) and 462 TUs (65.5%).

Of the 186 second-year medical students who began the two academic years, 178 (95.7%) completed all CCEs and TBL modules (86 men, 92 women, mean age 25.3 years). We analyzed the performance of 91 students in academic year 2003–2004 and 87 students in academic year 2004–2005, yielding 62,715 unique data points [(91 students × 345 questions) + (87 students × 360 questions)]. Scores for the 178 students included in this study are summarized in Table 3.

For both years combined, 178 students correctly answered 83.6% (SD 6.1) of TRs and 77.7% (SD 6.9) of TUs, achieving mean scores 5.9% (SD 5.5) higher on TRs ($P < .001$, *t* test) (Table 3). Similar results were observed when analyzing subgroups of term 1 or term 2 PBQs. For term 1 PBQs, students scored 4.8% (SD 7.0) higher on TRs than TUs ($P < .001$). A somewhat greater difference was observed for term 2, as students achieved 7.0% (SD 6.9) higher scores on TRs than TUs ($P < .001$). The mean discrimination index of TUs was slightly higher than TRs: 0.22 (TU) versus 0.20 (TR). This small difference is not surprising,

Table 3

Comparison of the Performance of 178 Second-Year Medical Students on Pathology-Based Exam Questions (PBQs), Boonshoft School of Medicine, 2003–2005*

| Group of PBQs | No. of questions | DI: Mean (SD) | Score | | |
|--------------------|------------------|---------------|-------------|------------|-----------------------------|
| | | | Mean % (SD) | Range % | <i>P</i> value [†] |
| All CCEs | | | | | |
| TR | 243 | 0.20 (0.12) | 83.6 (6.1) | 64.0–96.1 | <.001 |
| TU | 462 | 0.22 (0.13) | 77.7 (6.9) | 59.7–91.3 | |
| Term 1 CCEs | | | | | |
| TR | 127 | 0.20 (0.12) | 82.3 (7.3) | 59.7–98.5 | <.001 |
| TU | 267 | 0.22 (0.14) | 77.5 (7.2) | 60.0–93.6 | |
| Term 2 CCEs | | | | | |
| TR | 116 | 0.20 (0.13) | 85.0 (7.0) | 51.9–100.0 | <.001 |
| TU | 195 | 0.22 (0.13) | 78.0 (7.7) | 57.3–96.2 | |

* CCE indicates comprehensive course examination; DI, discrimination index; TR, TBL-related PBQ; TU, TBL-unrelated PBQ.

[†] The *P* value compares TR versus TU scores.

Table 4

Performance of Second-Year Medical Students in the Highest Academic Quartile (n = 45) Versus Those in the Lowest Academic Quartile (n = 45) on Pathology-Based Examination Questions (PBQs), Boonshoft School of Medicine, 2003–2005*

| Academic quartile and group of PBQ | Score on all exams | | Difference in scores [†] | |
|------------------------------------|--------------------|--------------|-----------------------------------|--------------|
| | Mean % (SD) | Range % | Mean % (SD) | Range % |
| Highest quartile | | | | |
| TR | 89.3 (4.0) | 80.6 to 96.1 | 3.8 (5.4) [‡] | –7.7 to 13.3 |
| TU | 85.5 (3.2) | 78.8 to 91.3 | | |
| Lowest quartile | | | | |
| TR | 77.5 (5.8) | 64.0 to 86.8 | 7.9 (6.0) [‡] | –5.1 to 20.6 |
| TU | 69.6 (4.5) | 59.7 to 77.5 | | |

* TBL, team-based learning; TR, TBL-related PBQ; TU, TBL-unrelated PBQ.

[†] TR versus TU scores.

[‡] $P = .001$ for two-way ANOVA interaction comparing the difference in mean scores on TR and TU questions for highest- versus lowest-quartile students.

considering that discrimination index is related to the difficulty of a test question, and students correctly answered TRs more often than TUs.

Analysis of students' performance by academic quartiles (Table 4) revealed that students in both the highest (n = 45) and lowest (n = 45) quartiles scored significantly higher on TRs compared with TUs ($P < .001$). Highest-quartile students achieved 3.8% (SD 5.4) higher scores on TRs, whereas lowest-quartile students scored 7.9% (SD 6.0) higher on TRs. Thus, the magnitude of the difference between TR and TU scores was greater in the lowest quartile compared with the highest quartile ($P = .001$, two-way ANOVA interaction).

Discussion and Conclusions

To the best of our knowledge, this is the first study in medical education demonstrating that TBL provides a larger learning benefit for lower-achieving students compared with higher-achieving students. Nevertheless, higher-achieving students also showed improved performance on comprehensive examinations, probably due to a combination of thorough study of the advance assignment and enhancement of personal knowledge through interaction with peers and faculty. Overall, students in two consecutive second-year classes demonstrated significantly higher performance on PBQs related to course content learned via TBL modules. In our opinion, a 5.9% higher mean score is large enough to be meaningful for educators and learners whose common goal is

achievement of learning objectives. We believe that these outcomes are especially encouraging to faculty who are considering TBL but are concerned about mastery of course content. Students' improved performance across the board, and particularly among the lowest-quartile performers, may reduce the failure rate on criterion-based examinations.

Our findings support both our hypotheses, suggesting that TBL has a positive impact on students' learning. Knowing that the IRAT will be administered at the beginning of a TBL session motivates students to prepare well by attempting to independently master knowledge contained in the advance assignment. Gaps and deficiencies in understanding are improved as peers explain to their teammates why they favor specific answers to questions as the group works toward consensus for the GRAT. Revealing all groups' answers simultaneously allows faculty to see which questions were not answered correctly by all teams. Faculty are then able to direct the ensuing discussion toward clarifying any difficult concepts that the GRAT showed were not well understood or were mastered incompletely. The culminating application exercises challenge each team to use their aggregate knowledge as they wrestle with faculty-designed problems. Teams must analyze information and negotiate to achieve consensus within a short time. After teams reveal their decisions, the intergroup discussion requires teams to explain to the class the evidence and reasoning that support their conclusions. As teams perceive how their

conclusions compare with others', faculty may further explore and extend the interpretations verbalized by learners.

TBL's sequential strategy motivates learners to go beyond mere mastery of essential facts. A well-crafted application exercise requires teams to apply knowledge to realistic situations, such as deciding which pathogenesis, diagnosis, or treatment is most likely or most appropriate for a particular patient. The process of arriving at consensus demands that students develop and demonstrate listening, teaching, and vigorous negotiation skills. The interteam discussion that follows provides every team with immediate comparative feedback regarding its conclusions. By deliberating over best answers within teams, and defending those answers to peers and faculty, students become engaged in learning why a particular choice is most appropriate. In describing the kinds of activities that enhance long-term learning, Frank Smith^{16(p87)} argues that "we can only learn from activities that are interesting and comprehensible to us; in other words, activities that are satisfying. If this is not the case, only inefficient rote learning, or memorization, is available to us and forgetting is inevitable." Medical students' higher performance on examination questions related to course content learned with the benefit of a TBL module suggests that TBL enhances mastery and retention of course content, at least over the duration of a single course.

The larger beneficial effect on examination performance for lowest-quartile students compared with highest-quartile students correlates with TBL's strategy. Pointed exchange among peers during the GRAT and application exercises, combined with faculty management of the interteam discussions, may be viewed as an orchestrated learning laboratory that helps students achieve a baseline of knowledge. Because peers are teaching each other while arriving at consensus answers, it seems reasonable that learning gains are likely to be greater for those who have less content mastery at the start of a TBL session. We have observed that no burdensome duty is placed on higher-performing students who begin the readiness assurance phase with a better

grasp of the advance assignment. Well-prepared students clarify their own knowledge by verbalizing and negotiating with peers, are rewarded with grades for their individual and team efforts, and spend no additional time accomplishing these tasks beyond the live session. However, students who arrive less prepared are not just enriched by their teammates' knowledge and critical thinking skills. They are also motivated by two factors to prepare more thoroughly for future sessions: the desire to achieve a better grade on the IRAT, and their peers' expectations that they will make valuable contributions to intrateam discussions. Peer influence, expressed through intragroup teaching and social pressure to prepare well, assists the academically challenged student in mastering course content.¹⁷

Effective implementation of TBL enables students who are academically "at risk" to learn significant portions of course content *before* CCEs, resulting in improved performance on those examinations. Other educational strategies, such as peer tutoring, have been shown in the literature to improve academic performance in health professions education.¹⁸ Peer tutoring, however, requires significant time commitments outside class. TBL benefits the at-risk student within the confines of class time, using the combined efforts of faculty and peers to promote learning.

Several limitations of this study's design and conclusions are apparent. First, because two authors' (P.K. and S.N.) content expertise is limited to pathology, we did not analyze performance on examination questions unrelated to pathology. Accordingly, our conclusions about TBL-related learning benefits may not apply to other medical science disciplines. Second, the argument could be made that any type of active educational strategy might enhance students' performance on examinations. Our findings may represent only the effect of dedicated class time rather than benefits of the TBL strategy. A prospective study comparing TBL with another active learning method in a single cohort of students is required to address that argument. Third, we recognize that using one person to categorize questions as TBL related may have introduced an unmeasured error. Fourth, we must consider the effect of

excluding 8/186 students (4.3% of the sample) because of incomplete exam data. Of these 8 students, 3 had acceptable academic standing and 5 were failing (the cumulative exam average of the latter group was <70%). We doubt that excluding only 4.3% significantly affects the whole group's mean TR and TU scores. However, exclusion of the 5 failing students (2.7% of the sample) alters the composition of the lowest quartile more than the highest quartile of students, so our data comparing quartile performances may be biased by their exclusion. Finally, and perhaps the most important limitation, examination performance was measured within four weeks after content-related TBL modules. Therefore, our results show benefits for relatively short-term learning.

Another concern is that systematic differences between TRs and TUs may have influenced results. Our design relies on difficulty value as the outcome measurement; that is, mean scores are equivalent to mean difficulty value. Therefore, we considered factors other than difficulty value to compare TRs with TUs. The difficulty of a multiple-choice question may be affected by structural features; a poorly written question introduces "artificial difficulty" that may affect students' performance.¹⁹ Structural features include format, wording, complexity of the stem, and the number of distracters. Two observations are pertinent to address this concern. First, we analyzed large numbers of questions in each group (TU and TR), increasing the probability that a similar range of formats was included in each group. Second, because all of the TUs and TRs that we analyzed were written or edited by two of this study's authors (P.K. and S.N.), structural features are likely to be similar in both groups. We also compared questions using the discrimination index, a useful measurement of item quality that reflects the degree to which a single test question differentiates between groups of students who scored well on the entire exam and those who scored poorly. The mean discrimination indices were 0.20 for TR and 0.22 for TU questions, indicating similar effectiveness in differentiating between the highest and lowest quartiles of students. These observations and results suggest that systematic differences in TRs versus TUs are unlikely.

More outcome-centered studies of TBL are needed to provide objective evidence of this active learning strategy's effectiveness in medical education. Potential benefits for longer-term learning need to be evaluated, such as performance on examinations administered several months after a TBL module or performance on comprehensive examinations assessing knowledge gained from several courses in which TBL was used. Additionally, a prospective research design that compares learning outcomes of academically similar student cohorts exposed to the TBL strategy versus another active learning method could produce meaningful data.

Acknowledgments: The authors thank Ms. Ife Shafeek, administrative coordinator of the Boonshoft School of Medicine's pathology department, for managing TBL materials and grade data with skill and grace. They are also indebted to Ms. Ruth Paterson and the academic affairs staff who provided complete comprehensive course examination data. They are grateful for the privilege of teaching and learning with the highly motivated medical students in the classes of 2006 and 2007.

Funding/Support: None.

Other disclosures: None.

Ethical approval: This study was deemed exempt by the Wright State University institutional review board.

References

- 1 Rawson RE, Quinlan KM. Evaluation of a computer-based approach to teaching acid/base physiology. *Adv Physiol Educ.* 2002;26:85–97.
- 2 Temkin B, Acosta E, Malvankar A, Vaidyanath S. An interactive three-dimensional virtual body structures system for anatomical training over the internet. *Clin Anat.* 2006;19:267–272.
- 3 Neville AJ. Problem-based learning and medical education forty years on. A review of its effects on knowledge and clinical performance. *Med Princ Pract.* 2009;18:1–9.
- 4 Neville AJ, Norman GR. PBL in the undergraduate MD program at McMaster University: Three iterations in three decades. *Acad Med.* 2007;82:370–374.
- 5 Enarson C, Cariaga-Lo L. Influence of curriculum type on student performance in the United States Medical Licensing Examination Step 1 and Step 2 exams: Problem-based learning vs. lecture-based curriculum. *Med Educ.* 2001;35:1050–1055.
- 6 Seidel CL, Richards BF. Application of team learning in a medical physiology course. *Acad Med.* 2001;76:533–534.
- 7 Haidet P, O'Malley KJ, Richards BF. An initial experience with team learning in medical education. *Acad Med.* 2002;77:40–44.

- 8 Michaelsen LK, Black RH. Building learning teams: The key to harnessing the power of small groups in higher education. In: Kadel S, Keener J, eds. Collaborative Learning: A Sourcebook for Higher Education. Vol 2. State College, Pa: National Center for Teaching, Learning, and Assessment; 1994.
- 9 Michaelsen LK, Knight AB, Fink LD, eds. Team-Based Learning: A Transformative Use of Small Groups in College Teaching. Sterling, Va: Stylus Publishing; 2004.
- 10 Michaelsen LK, Sweet M. Creating effective team assignments. In: Michaelsen LK, Parmelee DX, McMahon KK, Levine RE, eds. Team-Based Learning for Health Professions Education. Sterling, Va: Stylus Publishing; 2008.
- 11 Nieder GL, Parmelee DX, Stolfi A, Hudes PD. Team-based learning in a medical gross anatomy and embryology course. *Clin Anat*. 2005;18:56–63.
- 12 Levine RE, O’Boyle M, Haidet P, et al. Transforming a clinical clerkship with team learning. *Teach Learn Med*. 2004;16:270–275.
- 13 Letassy NA, Fugate SE, Medina MS, Stroup JS, Britton ML. Using team-based learning in an endocrine module taught across two campuses. *Am J Pharm Educ*. 2008;72:1–6.
- 14 Koles P, Nelson S, Stolfi A, Parmelee D, DeStephen D. Active learning in a year 2 pathology curriculum. *Med Educ*. 2005;39:1045–1055.
- 15 Nelson S, Koles P, Bierke-Nelson D. Down Syndrome: A Multidisciplinary Interactive Team-Based Learning Exercise. Available at: www.aamc.org/mededportal. ID 248. Accessed August 9, 2010.
- 16 Smith F. *The Book of Learning and Forgetting*. New York, NY: Teacher’s College Press; 1998.
- 17 Michaelsen LK, Sweet M. Fundamental principles and practices of team-based learning. In: Michaelsen LK, Parmelee DX, McMahon KK, Levine RE, eds. *Team-Based Learning for Health Professions Education*. Sterling, Va: Stylus Publishing; 2008.
- 18 Santee J, Garavalia L. Peer tutoring programs in health professions schools. *Am J Pharm Educ*. 2006;70:1–10.
- 19 Collins J. Writing multiple-choice questions for continuing medical education activities and self-assessment modules. *Radiographics*. 2006;26:543–551.