Acad Psychiatry (2016) 40:755–760
DOI 10.1007/s40596-016-0501-7

EMPIRICAL REPORT

A Randomized Controlled Trial of Team-Based Learning Versus Lectures with Break-Out Groups on Knowledge Retention

Grace C. Thrall[1] · **John H. Coverdale**[2] · **Sophiya Benjamin**[3] · **Anna Wiggins**[4] · **Christianne Joy Lane**[4] · **Michele T. Pato**[4,5]

Received: 19 April 2015 / Accepted: 1 February 2016 / Published online: 16 February 2016

Abstract
Objective This goal of this study was to evaluate the efficacy of team-based learning (TBL) on knowledge retention compared to traditional lectures with small break-out group discussion (teaching as usual (TAU)) using a randomized controlled trial.
Methods This randomized controlled trial was conducted during a daylong conference for psychiatric educators on attention-deficit hyperactivity disorder and the research literacy topic of efficacy versus effectiveness trials. Learners ($n=115$) were randomized with concealed allocation to either TBL or TAU. Knowledge was measured prior to the intervention, immediately afterward, and 2 months later via multiple-choice tests. Participants were necessarily unblinded. Data enterers, data analysts, and investigators were blinded to group assignment in data analysis. Per-protocol analyses of test scores were performed using change in knowledge from baseline. The primary endpoint was test scores at 2 months.
Results At baseline, there were no statistically significant differences between groups in pre-test knowledge. At immediate post-test, both TBL and TAU groups showed improved knowledge scores compared with their baseline scores. The TBL group performed better statistically on the immediate post-test than the TAU group (Cohen's $d=0.73$; $p<0.001$), although the differences in knowledge scores were not educationally meaningful, averaging just one additional test question correct (out of 15). On the 2-month remote post-test, there were no group differences in knowledge retention among the 42 % of participants who returned the 2-month test.
Conclusions Both TBL and TAU learners acquired new knowledge at the end of the intervention and retained knowledge over 2 months. At the end of the intervention day and after 2 months, knowledge test scores were not meaningfully different between TBL and TAU completers. In conclusion, this study failed to demonstrate the superiority of TBL over TAU on the primary outcome of knowledge retention at 2 months post-intervention.

Keywords Team-based learning · Lecture-based learning · Medical education · Evidence-based medicine · Residency training

Educators have shown a growing interest in active learning techniques to supplement or replace more passive instruction such as traditional lecture-based teaching. Problem-based learning, case-based group discussion, and other small-group learning formats have become popular in medical education. Over the past decade, team-based learning (TBL) has gained popularity as an active instructional strategy for medical educators. TBL is a disciplined and systematically constructed active learning modality for instructing large groups [1, 2]. As a consequence, educators are especially interested in learning about the effectiveness of TBL.

Two systematic reviews [3, 4] reported a dearth of randomized controlled trials that support the implementation of TBL compared with other learning strategies. While multiple observational cohort studies have found that TBL improves

✉ Grace C. Thrall
grace.thrall@duke.edu

[1] Central Regional Hospital, Butner, NC, USA
[2] Baylor College of Medicine, Houston, TX, USA
[3] McMaster University, Hamilton, ON, Canada
[4] Keck School of Medicine of University of Southern California, Los Angeles, CA, USA
[5] SUNY Downstate, Brooklyn, NY, USA

learner satisfaction and classroom engagement, fewer studies have addressed the question of whether or not TBL enhances knowledge retention. Among those studies measuring knowledge outcomes, there are only two published studies to our knowledge that have been conducted using a randomized controlled experimental design. Randomized controlled designs are regarded as the gold standard in education research because they permit inferences about causation, such as the conclusion that a particular instructional method causes better learning outcomes [5]. Our own search identified just two RCTs [6, 7]. In the first, Haidet compared TBL to standard slide and lecture teaching with 82 residents on knowledge of diagnostic testing 1 month after a single teaching session and found no differences between groups on knowledge outcomes [6]. As a significant limitation of this study, the balance of prognostic factors between groups created by randomization may have become unbalanced by 24 crossovers and 26 no-shows. In the second RCT, Yang compared TBL to lectures with 127 Chinese medical students in a 2-week neurology clerkship and found no differences on knowledge scores [7]. Other than RCTs, there are two prospective crossover studies of TBL, whose designs are limited by carryover, order, and contamination effects [8, 9]. Across all prospective observational studies, TBL appears to improve knowledge, but the results are mixed [4].

Because of the paucity and the generally low methodological quality of the available data, we set out to test the efficacy of TBL in a rigorous fashion using a randomized controlled design to assist educators in their judgments about the merits of TBL as a strategy for teaching. We conducted an RCT of TBL versus lecture with small-group discussion (or teaching as usual, TAU) in the setting of a daylong conference for psychiatry program directors on attention-deficit hyperactivity disorder (ADHD). Our primary goal was to determine whether TBL is more effective than TAU as a teaching method for content relevant to ADHD across the lifespan. We sought to determine whether or not TBL would increase knowledge retention compared to TAU at two time points, both immediately after the intervention and 2 months later. The primary endpoint was change from baseline test scores at 2 months.

Accordingly, we planned statistical analyses to answer the following questions: Are the remote post-test scores different based on group assignment after controlling for pre-test scores? Are the immediate post-test scores different based on group assignment after controlling for pre-test scores? Are the remote scores different after controlling for immediate post-test scores?

Methods

All who attended the daylong workshop on "Treating ADHD Across the Lifecycle: What to Learn and How to Teach It" for the 2010 American Association of Directors of Psychiatric Residency Training Annual Meeting were included in the study. There were no exclusion criteria. The University of Southern California provided IRB approval. Informed consent was obtained from all participants.

The workshop's three specific learning objectives were for participants to be able to describe the strengths and limitations of observational and experimental study designs, to differentiate efficacy trials from effectiveness trials, and to summarize the research evidence for treatment outcomes in ADHD. Our dual intentions were to improve both content knowledge and research literacy skill. To this end, faculty for TBL and TAU groups worked together to develop common learning objectives with enough specificity to permit each faculty group to independently create tightly linked learning activities. We wrote complex and challenging test questions to avoid a "ceiling effect" of perfect test scores and to measure growth in analytic skills essential to evidence-based practice, rather than content knowledge alone.

There were two distinct educational interventions, TBL and TAU. TAU consisted of traditional large group lectures with smaller-group break-out sessions, similar to the conference's customary format [10]. Two invited experts on ADHD gave three large group lectures totaling 3.5 h. There were two 45-min break-out sessions in which six small groups of seven to ten participants were led by seven facilitators. The break-out activities included a 15-min mini-talk by one group facilitator, 20 min of journal club-style review of an article using critical appraisal worksheets [11], and 10 min of facilitator-led group discussion. The same set of articles was used by both TAU and TBL groups. Additionally, the TAU break-out groups discussed an article on parent training for children with ADHD.

TBL was designed to include as many of the established, critical elements of the pedagogy as possible within the limits of a single-day intervention. Three faculty (SB, JC, GT) with extensive experience in TBL facilitated three TBL sessions. The TBL intervention incorporated most of the recommended TBL core design elements, including team formation, brief individual study, individual and group readiness assurance testing (iRAT and gRAT) with immediate feedback via "scratch-off" style answer cards (Immediate Feedback Assessment Technique (IF-AT) forms), and three intra-team "application" exercises. Application exercises incorporated each of the prescribed "Four Ss" [2] as follows: (1) a significant issue involving a patient-important problem in ADHD, (2) work on the same problem by all learners in each TBL group, (3) specific choice from among a list of instructor-provided options, and (4) simultaneous reporting by all learners using large cards displaying each group's decision. Although the TBL method calls for resources to be distributed equally across teams, our TBL groups were self-formed without attempt to distribute learners by ADHD knowledge,

evidence-based medicine knowledge, or TBL experience. Also, our single-day experience did not include a grading mechanism, peer feedback, or the opportunity for groups to develop into cohesive, well-functioning teams.

Participants randomized to the TBL group formed themselves into eight groups of approximately seven people each. Three TBL sessions of 70–90 min followed the standard TBL learning sequence: (1) a short period of self-study of provided learning materials; (2) iRATs and gRATs with moderated discussion; (3) team work on the application exercise; and (4) reporting of team answers, inter-team debate, and large group discussion. Each TBL session wrapped up with a brief summary of key learning points by the facilitators (learning materials for all three TBL sessions are available upon request).

All conference participants were randomized to either the TBL or TAU intervention. A computer-generated list of random numbers allocated subjects in a 1:1 ratio to each arm in fixed blocks of 10. The allocation sequence was concealed from researchers in sequentially numbered, opaque, sealed envelopes by office staff at the Department of Psychiatry & Behavioral Sciences at Duke University who were independent from the study. At the conference, each participant received an envelope with their group assignment from a staff person who was unaware of the allocation sequence. Participants wore a badge displaying both her name and group assignment for the entire day. Participants were necessarily unblinded to group assignment. Data enterers, data analysts, and investigators were blinded to subjects' group assignment.

We assessed group differences in knowledge retention at two time points—both immediately after the learning intervention and 2 months later. Knowledge retention was measured by internally developed multiple-choice tests written by two of the authors (GT, SB) that were designed to test content related to efficacy and effectiveness studies of ADHD. We obtained construct validity for the test questions by piloting them with both trainee and expert groups. At the conference, we administered the same pre-test to all participants at the beginning of the day. An immediate post-test was administered at the end of the day. A remote post-test was mailed to all participants 2 months later; a second copy of the same remote post-test was mailed 1 month after that to enhance response rates. All members of each group received the same test at each sitting. All three tests included 15 multiple-choice questions. To create the pre-test, immediate post-test, and remote post-tests, we pulled questions on each learning objective randomly from a common question bank to ensure that the topics and difficulty of each exam were equal. All test questions were mapped to the prespecified learning objectives. Sixty percent of the questions were identical on all three tests, but the order of the answer choices varied in order to reduce practice effects. The other 40 % of questions were unique. The authors anticipated that blinded randomization of questions would avoid the potential for bias and create equivalency among the three tests, although this was not verified.

In addition to knowledge outcomes, we also measured learner satisfaction via ratings on the question "I was satisfied with my experience today," learner efficiency via ratings on the question "I experienced an optimal amount of new learning for my time and effort in today's activities," and group engagement via the nine-item Student Self-Report of Engagement Measure [12].

Descriptive statistics (M, SD, 95 % CI) were computed for raw scores, change scores, and percent change for knowledge scores. Group differences in raw knowledge scores at each collection were tested using independent t tests. Comparisons of each group's change in scores from one data collection point to another were tested ANCOVA model with difference score as the outcome, group as the factor, and pre-test as the covariate. Additional exploratory models were performed post hoc on immediate change to determine if being in the lowest quartile of scores had an impact on scores.

Results

There were 155 conference participants (124 program directors, 2 residents, and 29 psychiatric educators and interested others) who registered in advance for the conference. Of those registered, 115 attended and were randomized, 56 to the TBL group and 59 to the TAU group. All randomized subjects completed the pre-test. In the TBL group, 51 (91 %) completed the immediate post-test and 24 (43 %) completed the remote post-test. In the TAU group, 55 (93 %) completed the immediate post-test and 24 (41 %) completed the remote post-test. Overall, 42 % of all randomized participants completed the remote post-test.

At baseline, there were no statistically significant differences in knowledge between groups. Learners in both TBL and TAU groups improved in knowledge on the immediate post-test compared to their baseline scores. On the end-of-the-day immediate post-test, there was a statistically significant difference in knowledge scores between groups (Table 1). Group assignment was a significant predictor of immediate post change, adjusting for pre-test ($p = 0.001$). This between-group difference in change at immediate post-test was large (effect size Cohen's $d = 0.7$), although this difference translated, on average, to one more question correct in the TBL group compared with the TAU group. Both groups sustained their knowledge improvements on the remote post-test (Fig. 1). There were no statistically significant differences between groups on knowledge retention at 2 months ($p = 0.67$). Additionally, there were no significant differences between groups at the end of the workshop day on learner satisfaction ($p = 0.93$) or learning efficiency ($p = 0.34$). On measures of group engagement, TBL learners showed significantly greater

Table 1 Knowledge test scores of team-based learning (TBL) vs. teaching as usual (TAU) groups

	TBL			TAU						
	N	*M* (95 % CI)	SD	*N*	*M* (95 % CI)	SD	*t*	*df*	*p*	*d*
Pre-test	56	9 (8.1, 9.8)	2.6	59	8.1 (6.7, 9.5)	2.8	0.78	113	0.44	0.15
Immediate post-test	51	11.7 (11.1, 12.3)	1.9	55	9.7 (9, 10.4)	2.3	3.73	104	<0.001	0.73
Remote post-test	24	11 (10.2, 11.7)	1.7	24	10.7 (9.7, 11.7)	2.3	0.43	46	0.67	0.12

TBL team-based learning, *TAU* teaching as usual with traditional lectures and small break-out group discussion, *N* number per group, *M* mean correct answers out of 15 possible on knowledge test, *SD* standard deviation, *t* *t* test, *p* significance test, *D* Cohen's *d*

engagement on eight of nine questions compared with TAU learners.

Given the study's high loss to follow-up, we compared the characteristics of those who completed the 2-month post-test versus those who did not. Similar numbers of learners were lost to follow-up from each group. We did not find a significant effect of either pre-test knowledge or immediate post-test knowledge (adjusted for group) on the likelihood of completing the remote post-test. There was a trend for participants who had higher learner satisfaction scores to return the remote survey ($F_{1,102}=3.0$, $p=0.08$), an effect that was not seen with reported learning efficiency ($p=0.24$).

We further explored the question of whether TBL was more efficacious for participants with lower baseline scores. Learners in the lowest quartile at baseline improved the same amount on the end-of-day immediate post-test, regardless of group assignment. A two-way ANOVA examining interaction of group and quartile effect was nonsignificant ($p=0.44$).

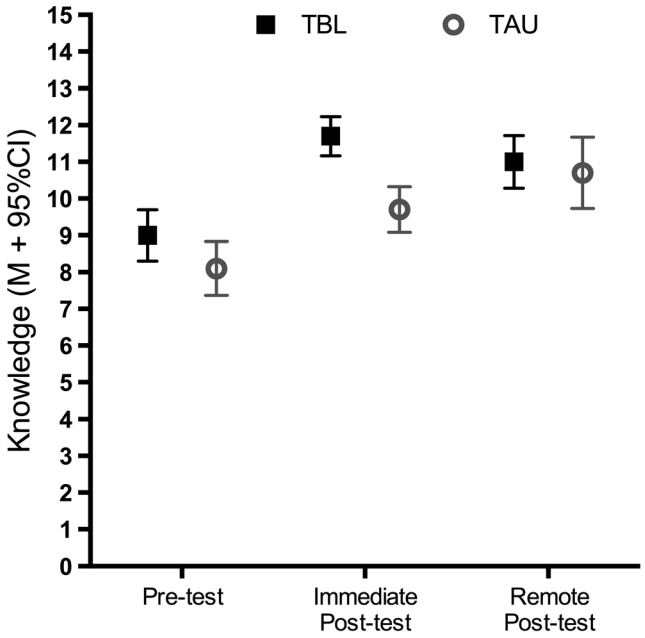

Fig. 1 Knowledge test scores of team-based learning (TBL) vs. teaching as usual (TAU) groups with 95 % confidence intervals at three time points. Figure shows Table 1 mean knowledge scores by group on a 15-item multiple-choice test administered at the start of the conference day (pre-test), end of day (immediate post-test), and 2 months later (remote post-test). *TBL* team-based learning, *TAU* teaching as usual with traditional lectures and small break-out group discussion

Discussion

Both TBL and teaching as usual (TAU) lecture-discussion groups improved in knowledge of concepts of efficacy and effectiveness related to attention-deficit hyperactivity disorder (ADHD) between the pre-test and both follow-up tests. The TBL group performed better statistically on the immediate post-test than the TAU group (Cohen's $d=0.73$; $p<0.001$), although the differences in knowledge scores were not educationally meaningful, averaging just one additional test question correct (out of 15). Moreover, we found no significant differences between groups 2 months after the educational intervention based on the scores of the 42 % of participants who returned the final test.

This study failed to demonstrate the superiority of TBL over TAU as hypothesized on the primary outcome of knowledge retention at 2 months post-intervention. The failure to demonstrate an advantage for TBL at the 2-month follow-up was limited by a substantial dropout rate.

These data should be understood in the context of prior findings of TBL effectiveness. Of two previously conducted prospective crossover studies of TBL, one found no significant differences in knowledge performance between TBL and case-based group discussion [8]. The second found that the performance of students taught by TBL was better on two of three topics than that of students taught by small group lectures [9]. Similarly, only 7 of 14 included studies in the most recently published systematic review on the effectiveness of TBL in health professions education showed that TBL appeared to improve knowledge scores [4]. Moreover, our data are consistent with the only two previously conducted RCTs of TBL [6, 7] showing no difference in knowledge retention between groups.

Some of the strengths of this study include the large sample size (115) of learners from various regions across the USA and a strong experimental design that minimized the potential for bias, including randomization with allocation concealment

and blinding of data analysts and investigators. Additionally, we believe that this study successfully demonstrated a "proof of concept" for an intensive 1-day implementation of team learning where the core elements of the model were employed with a large and diverse group of mostly TBL-naïve educators—a success that may encourage implementation at other workshop-style educational events.

The limitations of this study should be viewed in light of the feasibility challenges in educational research. First, the short length of this trial did not permit implementation of all elements of TBL, which would ideally incorporate features such as learner assignment to teams, longer pre-class preparation time period, peer evaluations and grading for accountability, and the time needed to develop highly functional teams. These limitations potentially reduced the impact of TBL on learning [1]. Secondly, written tests were primarily prepared by TBL instructors which may have advantaged TBL learners over TAU learners, who may or may not have received the same content emphasis by TAU small-group leaders. All instructors, however, were informed of the meeting's learning objectives in advance but not the specific content of application exercises. Thirdly, although subject retention was high for the immediate post-test, the high loss to follow-up (58 %) at 2 months created a large percentage of missing data on the study's primary outcome of knowledge retention. This high loss-to-follow-up substantially lowered statistical power to detect group differences on remote testing. While dropouts were approximately equal in each group, the low number of 2-month post-test completers creates the potential for bias in the results. For example, less confident or less expert subjects may have avoided returning the 2-month post-test, skewing the results toward an overestimate of TBL and/or TAU benefit. Lastly, the baseline characteristics of each group that may have had prognostic importance (such as prior knowledge of the ADHD literature, years as an educator, prior experience with TBL, or motivation to learn about TBL) were not measured as a practical choice to limit the information-gathering burden on participants. We expected, but could not confirm, that randomization would distribute these known prognostic factors equally between groups.

In conclusion, this rigorous experimental design did not find statistically significant differences in knowledge retention between TBL and lecture with small-group discussion at longer-term follow-up. There remains a dearth of experimental evidence for the efficacy of TBL on knowledge acquisition and retention in comparison with other active learning methods. Future research might strive to measure remote knowledge retention, which is a difficult but important and meaningful challenge in educational research. Furthermore, additional rigorous experimental designs are needed to assess the utility of teaching TBL in its "full-featured" format over extended time periods with a variety of learner groups in diverse settings.

Implications for Educators

- TBL can be used successfully with large conference groups that have traditionally depended on a lecture format and in circumstances where access to subject experts is limited.
- Because this randomized controlled trial of team-based learning compared to traditional lecture-discussion groups did not demonstrate an advantage of team-based learning, there remains insufficient RCT-level proof to recommend routine implementation of this intensive pedagogy.
- Measuring the impact of active learning strategies is vital for evidence-based curriculum design. The strengths and promise of team learning should encourage educators and education researchers to continue evaluating the efficacy of TBL on meaningful outcomes that relate to knowledge retention and lifelong learning skills.

Acknowledgments The authors wish to thank Paul Haidet for his contributions toward educating psychiatric educators about team-based learning.

This study was supported by the American Association of Directors of Psychiatry Residency Training (AADPRT) and funded by an R-13 grant from NIMH: MH074298 Teaching Scholarly Activity in Psychiatric Training 2005–2010.

Compliance with Ethical Standards

Disclosures On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Michaelsen LK, Parmelee DX, McMahon KK, Levine RE. Team-based learning for health professions education: a guide to using small groups for improving learning. Sterling: Stylus; 2004.
2. Haidet P, Levine RE, Parmelee DX, Crow S, Kennedy F, Kelly A, et al. Guidelines for reporting team-based learning activities in the medical and health sciences education literature. Acad Med. 2012;87(3):292–9.
3. Sisk RJ. Team-based learning: systematic research review. J Nurs Educ. 2011;50(12):665–9.
4. Fatmi M, Hartling L, Hillier T, Campbell S, Oswald AE. The effectiveness of team-based learning on learning outcomes in health professions education: BEME guide no. 30. Med Teach. 2013;35(12):e1608–24.
5. Mayer RE. The failure of educational research to impact educational practice: six obstacles to educational reform. In: Phye GD, Robinson DH, Levin J, editors. Empirical methods for evaluating educational interventions. San Diego: Elsevier Academic Press; 2005. p. 75.
6. Haidet P, Morgan RO, O'Malley K, Moran BJ, Richards BF. A controlled trial of active versus passive learning strategies in a large group setting. Adv Health Sci Educ. 2004;9(1):15–27.
7. Yang LH, Jiang LY, Xu B, Liu SQ, Liang YR, Ye JH, et al. Evaluating team-based, lecture-based and hybrid learning methods for neurology clerkship in China: a method-comparison study. BMC Med Educ. 2014;14:98.

8. Koles P, Nelson S, Stolfi A, Parmelee D, Destephen D. Active learning in a year 2 pathology curriculum. Med Educ. 2005;39(10):1045–55.
9. Thomas PA, Bowen CW. A controlled trial of team-based learning in an ambulatory medicine clerkship for medical students. Teach Learn Med. 2011;23(1):31–6.
10. Pato MT, Cyr RL, Manley LN, Morley CP. What to learn and how to teach it: five years of pre-meetings for training directors in psychiatry. Acad Psychiatry. 2013;37(2):76–81.
11. Pato MT, Cobb RT, Lusskin SI, Schardt C. Journal club for faculty or residents: a model for lifelong learning and maintenance of certification. Int Rev Psychiatry. 2013;25(3):276–83.
12. O'Malley KJ, Moran BJ, Haidet P, Seidel CL, Schneider V, Morgan RO, et al. Validation of an observation instrument for measuring student engagement in health professions settings. Eval Health Prof. 2003;26(1):86–103.