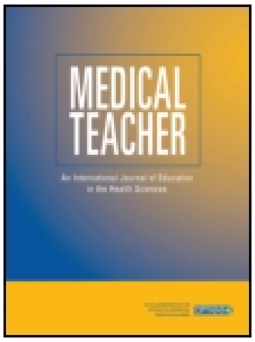

Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: https://www.tandfonline.com/loi/imte20

Effects of Team-Based Learning on short-term and long-term retention of factual knowledge

Amanda R. Emke, Andrew C. Butler & Douglas P. Larsen

To cite this article: Amanda R. Emke, Andrew C. Butler & Douglas P. Larsen (2016) Effects of Team-Based Learning on short-term and long-term retention of factual knowledge, Medical Teacher, 38:3, 306-311, DOI: 10.3109/0142159X.2015.1034663

To link to this article: https://doi.org/10.3109/0142159X.2015.1034663

Published online: 21 Apr 2015.

2016, 38: 306–311 MEDICAL TEACHER

Effects of Team-Based Learning on short-term and long-term retention of factual knowledge*

AMANDA R. EMKE[1], ANDREW C. BUTLER[2] & DOUGLAS P. LARSEN[1]
[1]Washington University School of Medicine, USA, [2]University of Texas, USA

Abstract

Purpose: We investigated the effect of Team-Based Learning (TBL) on long-term retention of knowledge in comparison to a traditional curriculum.

Methods: As TBL was incorporated into our curriculum in the 2008–2009 academic year, students were compared with those who received the traditional curriculum the year prior. Students in both the groups completed multiple-choice knowledge test at four time points spanning two years. Test performance was compared at each time point to assess changes in knowledge retention as a function of time.

Results: Baseline knowledge did not differ significantly between the TBL and control groups [51% versus 46%; $t(84) = 0.91$, $p = 0.37$, $d = 0.20$]. Performance improved after the course for both the groups, but was significantly higher in the TBL group [79% versus 59%; $t(84) = 4.96$, $p = 0.000004$, $d = 0.95$]. However, when assessed prior to the pediatrics clerkship, learning gains from TBL had largely disappeared and the small difference in performance was not significant [57% versus 51%; $t(84) = 1.51$, $p = 0.14$, $d = 0.32$].

Conclusion: Incorporating TBL into the pre-clinical pediatrics curriculum led to large gains in knowledge over the short-term, but these gains did not persist. Further research should focus on extending the impact of TBL on long-term knowledge retention.

Introduction

Medical students must acquire vast amount of knowledge during the pre-clinical portions of their undergraduate medical education (UME). Since this knowledge forms the foundation for the clinical training that follows, it is imperative that students retain it over time from pre-clinical courses to the corresponding clinical rotations, a period that can range from weeks to years. In our experience, students often show poor long-term retention of pediatric knowledge between their pre-clinical pediatrics course and their pediatrics clerkship. Correspondingly, studies have demonstrated that a substantial portion of the basic knowledge learned through medical education is forgotten over the long-term (Ling et al. 2008; Custers 2010, Custers & ten Cate 2011; Sullivan et al. 2013).

In order to achieve the goal of facilitating long-term retention of information (among other educational objectives), medical education has begun to move away from the traditional pedagogical method of the didactic lecture to incorporate active learning strategies (Spencer & Jordan 1999). Rather than treating students as passive receivers of knowledge, active learning engages them in higher-order cognitive tasks that enable them to construct their own knowledge in a meaningful way (Schmidt et al. 1989; Slavin 1996; van Blankenstein et al. 2011, 2013). Numerous studies in a broad range of disciplines and educational settings have demonstrated that active learning improves a variety of learning outcomes, including retention, critical thinking, communication and peer teaching (Gokhale 1995; Slavin 1996; Kvam 2000; Terenzini et al. 2001; Haidet et al. 2004; Prince 2004; Narli 2011). A recent meta-analysis on the benefits of active learning in science, technology, engineering and mathematics courses found an increase in examination results and a decrease in failure rates (Freeman et al. 2014).

One pedagogical method that facilitates active learning is Team-Based Learning (TBL). In TBL, each class session consists of a series of six learning activities that was build upon each other (pre-classroom work, individual assessment

Practice points

- Team-based learning improved the acquisition of knowledge.
- Team-based learning had little impact on long-term retention of knowledge.
- Continued practice through test-enhanced learning may improve TBL's impact.
- Medical education research should assess long-term outcomes.

Correspondence: Dr Amanda R. Emke, Department of Pediatrics, Division of Critical Care Medicine, St. Louis Children's Hospital, One Children's Place, NWT CB 8116, St. Louis, MO 63110, USA. Tel: 1 314 454 2527; Fax: 1 314 361 0733; E-mail: emke_a@kids.wustl.edu

*Research was conducted at Washington University School of Medicine under exempt and waived consent status.

ISSN 0142-159X print/ISSN 1466-187X online/16/030306–6 © 2015 Informa UK Ltd.
DOI: 10.3109/0142159X.2015.1034663

of content mastery, team assessment of content mastery, application exercises, appeals and peer feedback) to create increased student engagement (Kelly et al. 2005; Michaelsen & Sweet 2008). Numerous studies have demonstrated the efficacy of TBL in medical education through improved performance on end-of-course assessments and enhanced perceptions of learning, without impairing satisfaction ratings among students or faculty (McInerney & Fink 2003; Searle et al. 2003; Thompson et al. 2004; Nieder et al. 2005; Chung et al. 2009; Shellenberger et al. 2009; Wiener et al. 2009; Fatmi et al. 2013).

Despite the apparent benefits of TBL with respect to a host of outcome measures, it is unclear whether TBL improves the long-term retention of knowledge relative to traditional lecture-based pedagogical approaches. Educators often assess the impact of educational interventions soon after the completion of learning; however, in the absence of further practice, performance on such assessments is a poor predictor of long-term retention (Bjork 1994; Ling et al. 2008; Custers 2010; Custers & ten Cate 2011; Sullivan et al. 2013). Given the amount of work required to redevelop curricula to implement TBL, it is necessary to demonstrate that TBL improves both short-term and long-term knowledge retention. Our research investigates the impact of adding TBL to a lecture-based curriculum on the acquisition and long-term retention of pediatric knowledge. Given the findings of previous studies, we hypothesized that implementing TBL in a pre-clinical pediatrics course would improve performance on a knowledge test given immediately after the course. We also investigated whether any benefits of TBL would persist over longer time periods.

Methods

Using a quasi-experimental design, we assessed the impact of TBL on short-term and long-term retention of pediatric knowledge. TBL was incorporated into the pre-clinical pediatrics course at Washington University School of Medicine to create a TBL-plus-lecture curriculum. The student cohort in this new TBL-plus-lecture curriculum was compared to the student cohort from the prior year, who took the course with a lecture-only curriculum. For both the groups, pediatric knowledge was assessed at four-time points: (1) pre-course, (2) post-course, (3) pre-clerkship and (4) post-clerkship.

Subjects

In each year (2007–2008 and 2008–2009), 122 second-year students took the pre-clinical pediatrics course. All students had the opportunity to participate in the study; however, it was planned that students in the MD/PhD program and students who remediated the course would be excluded from analysis.

Course information

The pre-clinical pediatric course for students in the 2007–2008 academic year (control group) consisted of 16 one-h lecture sessions. Eight lectures covered core pediatric topics (e.g. newborn medicine, adolescent medicine, guidelines for the pediatric history and physical exam) while another eight lectures covered organ-specific pediatric topics (e.g. cerebral palsy, evaluation of the pediatric patient with wheezing, pediatric infectious disease). Student grades were based solely on a cumulative multiple-choice exam at the end of the course.

For students in the 2008–2009 academic year (intervention group), the course was modified to include both lecture and TBL sessions. Creating a TBL-plus-lecture curriculum involved changes to the content and structure of the course. The TBL-plus-lecture curriculum began with lectures on the core pediatric topics followed by TBL sessions that focused on the organ-specific topics but also built upon the lectures (Haidet et al. 2012). In total, the course contained 7 h of lecture (two core pediatric topics were condensed into a single lecture) and 10 h of TBL (the eight organ-specific topics were reorganized to form five two-hour sessions). Thus, the TBL-plus-lecture version of the course contained one extra hour of class time relative to the lecture-only version, but this difference largely reflected the addition of new organ-specific material which was not covered in the tests used in this study.

Each session was developed by a single instructor and facilitated by the developer plus one additional pediatric faculty member. Both faculty members were novice TBL instructors, but the developer had attended conferences and workshops prior to implementation. Students had been exposed to TBL in one previous course where it was used for review. The class of 122 students was divided into 18 teams by the Office of Medical Student Education with each team containing 6–7 students, and the teams were held constant over the two courses. The only rubric used in team formation was to ensure each team had one, but no more than two, MD/PhD students. Suggested resources for pre-classroom TBL preparation included sections from textbooks or review articles as well as review of previous lectures. TBL sessions were held in a large lecture hall beginning with an individual readiness assurance test (IRAT) – a 10-question multiple-choice test completed without the use of notes that covered all preparatory content as guided by the learning objectives. Next, students completed the group readiness assurance test (GRAT) by taking the identical multiple-choice test as a team, again without notes. No feedback was provided after the IRAT, but feedback was given after the GRAT using an audience response system, which allowed students to see the distribution of each answer choice on the GRAT. Each question was discussed to ensure thorough understanding of each concept.

Once understanding was assured, teams completed a series of clinically-based application exercises about the content identified in the learning objectives. Each application exercise consisted of 2–3 sections with multiple-choice questions at the end of each section. Teams were simultaneously reporting their answer to each question *via* the audience response system and then held up cards demonstrating the letter of their answer choice to facilitate across-team discussion. Appeals were allowed at the team level and awarded based on facilitator judgment. Student grades were based on TBL performance as scored by the multiple-choice questions on the IRAT, GRAT and application exercises as well as a cumulative, end-of-course multiple-choice exam that covered the lecture material.

Outcome measures

Knowledge retention was measured by a short multiple-choice test that targeted key educational objectives from the course (the test is available from the first author upon request). The main rationale for using a separate test from the criterial exam in the course was feasibility: it was unlikely that students would be willing to devote their free time to repeatedly taking a full course exam. The knowledge test only assessed material that was covered in both the courses. All of this material had been tested at some point during the course (i.e. the final exam, IRAT/GRAT, etc.) but using different questions. Thus, the questions on the knowledge test were novel to the students. The test was administered to students in both the intervention and control groups at four time points: (1) the beginning of the course (baseline), (2) the end of the course, (3) the beginning of the student's pediatric clerkship and (4) the end of clerkship. Scores on the test were calculated as the percentage of the questions answered correctly.

In addition to the knowledge questions, students were also asked a question about their preferences regarding different learning methods. In response to the statement "The method by which I prefer to learn is:", students could select "individual, self-study", "group study", "lecture", any combination of the three choices or "none of the above".

Statistical analysis

Performance on the knowledge tests was analyzed in several different ways. First, a MANOVA was conducted with group as an independent variable, the first knowledge test as a covariate, and the subsequent three knowledge tests as dependent measures. Second, a MANOVA was conducted with group as an independent variable and changes to scores representing gain from the first (baseline) knowledge test to the three subsequent tests as dependent measures. Third, the difference in performance between the two groups at each time point was assessed using independent *t*-tests. Since all three methods of analysis yielded the same pattern of results, we chose to present the *t*-test analysis because it is the simplest to interpret. Tests were performed using IBM SPSS Statistics version 21.0 (IBM Corp., Armonk, NY). Effect sizes were interpreted based on Cohen's recommendations (0.2–0.49 = small, 0.5–0.79 = medium, $\geq$0.8 = large) (Springer et al. 1999; Publication Manual of the American Psychological Association 2010).

Results

Subjects

Student demographics were collected for both academic years (Table 1). Only students with data for all four testing points were included for final analysis. As expected given the difficulty of tracking students over a multi-year longitudinal study, not all of the students who took the course completed all of the knowledge tests. In total, 82 students in the control group and 76 students in the intervention group missed one or more of the tests. Common causes for missing data included participation in the MD/PhD program, remediation or improper identification. Students with incomplete data were excluded from the analysis leaving a total of 40 students in the control group and 46 students in the intervention group.

Table 1. Demographics for all students who enrolled in the course as a function of cohort.

	Control group	Intervention group
Number of students	122	122
Male (%)	61 (50)	62 (51)
Caucasian (%)	63 (52)	63 (52)
GPA		
Mean	3.84	3.88
Median	3.89	3.92
Mode	4.00	3.98
Standard deviation	0.16	0.13
Range	3.14–4.00	3.28–4.00
MCAT scores		
Mean	37	37
Median	37	38
Mode	37	37
Standard deviation	2.77	2.67
Range	27–41	29–42

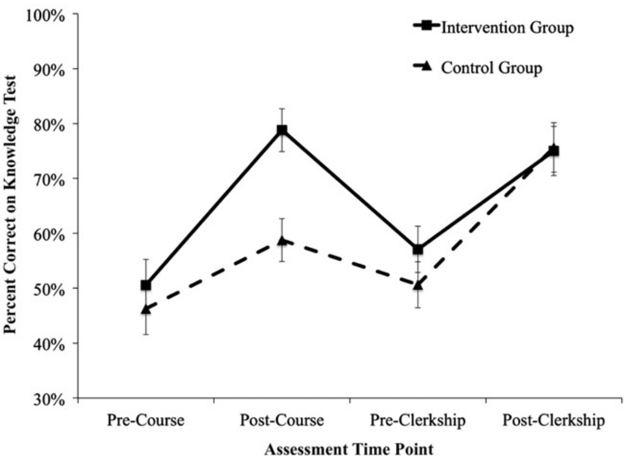

Figure 1. Percentage of correct responses on the knowledge test as function of group (intervention versus control) and assessment time point. Error bars represent 95% confidence intervals.

Test scores

Figure 1 depicts the percentage of correct responses on the knowledge test as function of group (intervention versus control) and assessment time point. Performance on the pre-course knowledge test was slightly higher in the intervention group relative to the control group, but this difference was not significant [51% versus 46%; $t(84) = 0.91$, SED = 4.72, $p = 0.37$, $d = 0.20$]. Thus, both the groups demonstrated comparable levels of knowledge prior to taking the course. As expected given that the material on the test was covered in the course, performance increased from the pre-course baseline to the post-course assessment in both groups (Figure 1); however, the gains in knowledge differed markedly between the two groups – the intervention group performed significantly better on the post-course knowledge test than the control group [79% versus 59%; $t(84) = 4.96$, SED = 4.04, $p = 0.000004$, $d = 0.95$].

Interestingly, the gains in knowledge that resulted from taking the second-year course were not sustained over the long-run. Performance on the third knowledge test demonstrated that substantial forgetting occurred in both the groups between the end of the course and the beginning of the third-year pediatric clerkship. The intervention group performed slightly better than the control group, but the difference was not significant [57% versus 51%; $t(84) = 1.51$, SED = 4.27, $p = 0.14$, $d = 0.32$]. When knowledge was assessed for the final time after students had completed their clerkship, performance had improved substantially and there was no significant difference between the intervention and control groups [75% versus 76%; $t(84) = 0.136$, SED = 4.61, $p = 0.89$, $d = 0.03$]. Presumably, the clerkship helped students to re-learn much of the material that they had forgotten prior to clerkship.

In order to rule out possible issues with interpreting these findings, we conducted a few follow-up analyses. First, we analyzed all the test data collected at each time point irrespective of whether or not students had completed all of the assessments. This analysis yielded the same pattern of results as that reported above, which indicates that the acquisition and retention of knowledge was similar in students who did and did not complete all of the assessments. Second, we explored whether student preferences about learning methods influenced the effects of the intervention. Students were divided into two groups based on their response to the learning preferences question at the beginning of the course: (1) those who preferred individual learning only (i.e. self-study, lecture or a combination of the two) and (2) those who preferred group study or both group and individual learning (i.e. group study only or in combination with other learning methods). A 2 (curriculum: intervention, control) × 2 (learning preference: individual-only, group) MANOVA was conducted with performance on the first knowledge test as a covariate and performance on the subsequent three tests as dependent variables. The analysis revealed no main effect or interaction of learning preference at any of the three time points (all Fs <1.6). Third, we examined whether the time between the completion of the second-year course and the beginning of clerkship influenced the performance on the pre-clerkship knowledge test. Students in both the groups went through their third-year clerkship at various times during the academic year with 8–40 weeks elapsing in between the end-of-course exam and beginning of the clerkship. An ANOVA was conducted with group as the independent variable, performance on the third knowledge test (pre-clerkship) as the dependent variable and both performance on the first knowledge test and the number of days before the start of clerkship as covariate. The main effect of number of days before clerkship was not significant ($F < 1$) and, as reported above, the main effect of group on performance on the pre-clerkship knowledge test was also not significant [$F(1,82) = 1.79$, MSE = 697.01, $p = 0.19$, $\eta_p^2 = 0.02$].

Discussion

Our research confirms the findings of previous studies that have shown robust benefits to learning from implementing TBL when knowledge is assessed during or at the end of the course (McInerney & Fink 2003; Kuhne-Eversmann et al. 2008; Chung et al. 2009; Wiener et al. 2009; Koles et al. 2010). However, our research also revealed a novel finding: the gains in knowledge that TBL produced did not persist over longer periods of time. That is, TBL failed to produce better long-term retention of knowledge relative to the traditional lecture format. This pattern of results highlights the dangers of assuming that performance on immediate assessments is predictive of long-term retention (Bjork 1994; Ling et al. 2008; Custers 2010, Custers & ten Cate 2011; Sullivan et al. 2013). The sizable gains in knowledge that result from dynamic learning methods, such as TBL, can be easily lost in the absence of continued practice.

The findings of our study are important in that they demonstrate the need for a renewed emphasis on assessing long-term outcomes in medical education to complement the ongoing movement toward more effective pedagogical methods. Over the past few decades, there has been a laudable shift towards the use of active learning methods. For example, the Liaison Committee on Medical Education (LCME), the accrediting body for medical education in the United States, has encouraged the increased use of active learning in UME (LCME Accreditation Standards 2014). TBL has become increasingly popular as an instructional method to incorporate active learning into pre-clinical medical education, but other methods exist as well. Taken as a whole, it is clear that these various methods of implementing active learning positively impact student achievement and the acquisition of knowledge (Benware & Deci 1984; Gokhale 1995; Slavin 1996; Johnson et al. 2000; Kvam 2000; Terenzini et al. 2001; Prince 2004). However, it remains unclear how well these various methods promote the long-term retention of knowledge because most studies do not assess long-term outcomes. Our findings, while specific to TBL, suggest that we must continue to critically evaluate the long-term effectiveness of these educational changes to ensure that these interventions are beneficial for our trainees.

To be clear, we are not suggesting that TBL or other active learning methods are ineffective. Clearly, the post-course gains in knowledge from TBL observed in our research and other studies are robust. TBL impacts learning within a course through the combination of a variety of mechanisms known to promote learning in isolation: it provides an opportunity for students to apply knowledge in real-world medical scenarios in a cooperative learning environment (Benware & Deci 1984; Gokhale 1995; Johnson et al. 2000; Terenzini et al. 2001; Haidet et al. 2004; Prince 2004; Kelly et al. 2005; Nieder et al. 2005; Bruno et al. 2007; Ling et al. 2008; Koles et al. 2010; Narli 2011; Sullivan et al. 2013); it creates an environment of increased student engagement while ensuring individual and team accountability by using peer instruction and discussion (Slavin 1996; Springer et al. 1999; Rao & DiCarlo 2000; Roa et al. 2002; Cortright et al. 2003; Prince 2004; Smith et al. 2009).

Rather, we argue that medical educators must take advantage of the robust gains in knowledge from TBL and other active learning methods by helping students to maintain these gains over time through continued practice. Recent research from cognitive psychology on test-enhanced learning provides a lens for how TBL could be modified to enhance long-term retention. Test-enhanced learning is an educational strategy that uses repeated retrieval of information (i.e. testing) to increase retention of knowledge (Larsen et al. 2008). In TBL, the IRAT, GRAT and application questions represent opportunities to practice retrieving and applying knowledge. However, such massed testing has limited benefits. For example, Schmidmaier et al. demonstrated that repeated tests back-to-back (similar to the format in TBL) led to significantly increased retention at one week, but the effect had disappeared by six months (Schmidmaier et al. 2011). In contrast, spacing out opportunities to practice retrieval over time generally leads to robust long-term retention as far out as six months and beyond (Larsen et al. 2009, 2013a,b). In the cognitive psychology literature, such findings are well-documented: for example, Cepeda et al. demonstrated that for knowledge to be retained on the order of months and years, repeated tests should be spaced by weeks and possibly even months (Cepeda et al. 2008; see too Rawson & Dunlosky 2011). Studies on test-enhanced learning suggest that the gains from TBL might be better maintained if it incorporated spaced testing, either within the course or after its conclusion. Future research will need to investigate this hypothesis further.

Limitations

Our study has several limitations, many of which result from the realities of conducting longitudinal classroom research. First, the short multiple-choice knowledge test used as the outcome measure only assesses a small portion of what students learned in the course. As explained above, this test was designed for practical reasons because there was a limit to the time students would devote to repeated testing. However, future research should try to expand the material covered by such a test. In addition, it would be interesting to assess higher-order thinking (e.g. applying knowledge to solve novel problems) in future research because such skills are a large part of what students learn through TBL. That said, basic knowledge is often a pre-requisite for higher-order thinking, so if knowledge is forgotten (as in the present study) it may impact higher-order thinking. Second, due to the limited amount of time people would engage to take the knowledge test, certain other variables (e.g. interest in the topic, motivation, etc.) that would be interesting to correlate were not collected. Finally, our study represents the effects of TBL in a single course. In order to confirm our findings, further studies on long-term retention should be carried out in other courses and at other institutions.

Conclusion

This study found that a lecture-plus-TBL pre-clinical pediatrics curriculum produced greater gain in pediatrics knowledge at the end of the course relative to a lecture-only curriculum. However, this effect did not persist when knowledge was assessed again after longer periods of time. Further research is needed to determine how to increase the long-term retention of the knowledge gained from TBL. Continued practice during or after the course through methods, such as test-enhanced learning, may be one way to achieve the goal of long-term retention. More broadly, the present findings suggest a need to assess long-term outcomes in studies on the efficacy of active learning in medical education.

Glossary

Test-enhanced learning: An educational strategy that uses repeated retrieval of information to increase retention of knowledge.

Larsen DP, Butler AC, Roediger HL. 2008. Test-enhanced learning in medical education. Med Educ 42:959–966.

Notes on contributors

Dr AMANDA R. EMKE, MD, is an Assistant Professor in the Division of Pediatric Critical Care Medicine, Department of Pediatrics.

Dr ANDREW C. BUTLER, PhD, is an Assistant Professor in the Department of Educational Psychology.

Dr DOUGLAS P. LARSEN, MD, MEd, is an Assistant Professor in the Division of Pediatric Neurology, Department of Neurology.

Acknowledgments

The authors would like to thank Susie Mueller and Liz Scott for their help in transcribing the data and administering the tests and surveys.

Declaration of interest: The authors report no declarations of interest.

This study was approved by the Washington University institutional review board under exempt and waived consent status.

References

American Psychological Association. 2010. Publication Manual of the American Psychological Association. 6th ed. Washington, DC: American Psychological Association.

Benware CA, Deci EL. 1984. Quality of learning with an active versus passive motivational set. Am Educ Res J 21:755–765.

Bjork RA. 1994. Institutional impediments to effective training. In: Druckman D, Bjork RA, editors. Learning, remembering, believing: Enhancing human performance. Washington DC: National Academy Press. pp 295–306.

Bruno P, Ongaro A, Fraser I. 2007. Long-term retention of material taught and examined in chiropractic curricula: Its relevance to education and clinical practice. J Can Chiropr Assoc 51:14–18.

Cepeda NJ, Vul E, Rohrer D, Wixted JT, Pashler H. 2008. Spacing effect in learning: A temporal ridgeline of optimal retention. Psychol Sci 19: 1095–1102.

Chung EK, Rhee JA, Baik YH, Oh-Sun A. 2009. The effect of team-based learning in medical ethics education. Med Teach 31:1013–1017.

Cortright RN, Collins HL, Rodenbaugh DW, DiCarlo SE. 2003. Student retention of course content is improved by collaborative-group testing. Adv Physiol Educ 27:102–108.

Custers EJ, ten Cate OT. 2011. Very long-term retention of basic science knowledge in doctors after graduation. Med Educ 45:422–430.

Custers EJ. 2010. Long-term retention of basic science knowledge: A review study. Adv Health Sci Educ 15:109–128.

Fatmi M, Hartling L, Hillier T, Campbell S, Oswald AE. 2013. The effectiveness of team-based learning on learning outcomes in health professions education: BEME Guide No. 30. Med Teach 35: e1608–e1624.

Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jodt H, Wenderoth MP. 2014. Active learning increases student performance in science, engineering, and mathematics. PNAS 111:8410–8415.

Gokhale AA. 1995. Collaborative learning enhances critical thinking. JTE 7:22–30.

Haidet P, Levine RE, Parmelee DX, Crow S, Kennedy F, Kelly PA, Perkowski L, Michaelsen L, Richards BF. 2012. Perspective: Guidelines for reporting team-based learning activities in the medical and health sciences education literature. Acad Med 87:292–299.

Haidet P, Morgan RO, O'Malley K, Moran BJ, Richards BF. 2004. A controlled trial of active versus passive learning strategies in a large group setting. Adv Health Sci Educ 9:15–27.

Johnson DW, Johnson RT, Stanne MB. 2000. Cooperative learning methods: A meta-analysis. [Accessed 8 August 2013] Available from: http://www.tablelearning.com/uploads/File/EXHIBIT-B.pdf.

Kelly PA, Haidet P, Schneider V, Searle N, Seidel CL, Richards BF. 2005. A comparison of in-class learner engagement across lecture, problem-based learning, and team learning using the STROBE classroom observation tool. Teach Learn Med 17:112–118.

Koles PG, Stolfi A, Borges NJ, Nelson S, Parmelee DX. 2010. The impact of team-based learning on medical students' academic performance. Acad Med 85:1739–1745.

Kuhne-Eversmann L, Eversmann T, Fischer MR. 2008. Team- and case-based learning to activate participants and enhance knowledge: An evaluation of seminars in Germany. J Cont Educ Health Prof 28: 165–171.

Kvam PH. 2000. The effect of active learning methods on student retention in engineering statistics. Am Stat 54:136–140.

Larsen DP, Butler AC, Roediger HL. 2008. Test-enhanced learning in medical education. Med Educ 42:959–966.

Larsen DP, Butler AC, Roediger HL. 2009. Repeated testing improves long-term retention relative to repeated study: A randomized controlled trial. Med Educ 43:1174–1181.

Larsen DP, Butler AC, Lawson AL, Roediger HL. 2013a. The importance of seeing the patient: Test-enhanced learning with standardized patients and written tests improves clinical application of knowledge. Adv Health Scie Educ 18:409–425.

Larsen DP, Butler AC, Roediger HL. 2013b. Comparative effects of test-enhanced learning and self-explanation on long-term retention. Med Educ 47:674–682.

Liaison Committee on Medical Education. Current LCME Accreditation Standards: Functions and structure of a medical school. [Accessed 6 February 2014] Available from: http://www.lcme.org/standard.htm.

Ling Y, Swanson DB, Holtzman K, Bucak SD. 2008. Retention of basic science information by senior medical students. Acad Med 83:S82–S85.

McInerney M, Fink LD. 2003. Team-based learning enhances long-term retention and critical thinking in an undergraduate microbial physiology course. JMBE 4:1–18.

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

Narli S. 2011. Is constructivist learning environment really effective on learning and long-term knowledge retention in mathematics? Example of the infinity concept. Educ Res Rev 6:36–49.

Nieder GL, Parmelee DX, Stolfi A, Hudes PD. 2005. Team-based learning in a medical gross anatomy and embryology course. Clin Anat 18:56–63.

Prince M. 2004. Does active learning work? A review of the research. J Eng Educ 93:223–231.

Rao SP, DiCarlo SE. 2000. Peer instruction improves performance on quizzes. Adv Physiol Educ 24:51–55.

Rawson KA, Dunlosky J. 2011. Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? J Exp Psychol-Gen 140(3):283–302.

Roa SP, Collins HL, DiCarlo SE. 2002. Collaborative testing enhances student learning. Adv Physiol Educ 26:37–41.

Schmidmaier R, Ebersbach R, Schiller M, Holzer M, Fischer MR. 2011. Using electronic flashcards to promote learning in medical students: Retesting versus restudying. Med Educ 45:1101–1110.

Schmidt HG, De Volder ML, DeGrave WS, Moust JH, Patel VL. 1989. Explanatory models in the processing of science test: The role of prior knowledge activation through small group discussion. J Educ Psychol 81:610–619.

Searle NS, Haidet P, Kelly PA, Schneider VF, Seidel CL, Richards BF. 2003. Team learning in medical education: Initial experiences at ten institutions. Acad Med 78:S55–S58.

Shellenberger S, Seale JP, Harris DL, Johnson JA, Dodrill CL, Velasquez MM. 2009. Applying team-based learning in primary care residency programs to increase patient alcohol screenings and brief interventions. Acad Med 84:340–346.

Slavin RE. 1996. Research on cooperative learning and achievement: What we know, what we need to know. Contemp Educ Psychol 21:43–69.

Smith MK, Wood WB, Adams WK, Wieman C, Knight JK, Guild N, Su TT. 2009. Why peer discussion improves student performance on in-class concept questions. Science 323:122–124.

Spencer JA, Jordan RK. 1999. Learner centered approaches in medical education. Br Med J 318:1280–1283.

Springer L, Stanne ME, Donovan SS. 1999. Effects of small-group learning on undergraduates in science, mathematics, engineering and technology: A meta-analysis. Rev Educ Res 69:21–51.

Sullivan PB, Gregg N, Adams E, Rodgers C, Hull J. 2013. How much of the paediatric core curriculum do medical students remember? Adv Health Sci Educ 3:365–373.

Terenzini PT, Cabrera AF, Colbeck CL, Parente JM, Bjorklund SA. 2001. Collaborative learning versus lecture/discussion: Students' reported learning gains. J Eng Educ 90:123–130.

Thompson BM, Schneider VF, Haidet P, Levine RE, McMahon KK, Perkowski LC, Richards BF. 2004. Team-based learning at ten medical schools: Two years later. Med Educ 41:250–257.

van Blankenstein FM, Dolmans D, van der Vleuten C, Schmidt HG. 2011. Which cognitive processes support learning during small-group discussion? The role of providing explanations and listening to others. Instr Sci 39:189–204.

van Blankenstein FM, Dolmans DH, Van der Vleuten CP, Schmidt HG. 2013. Relevant prior knowledge moderates the effect of elaboration during small group discussion on academic achievement. Instr Sci 41: 729–744.

Wiener H, Plass H, Marz R. 2009. Team-based learning in intensive course format for first-year medical students. Croat Med J 50:69–76.