Medical Teacher

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

The effectiveness of team-based learning on learning outcomes in health professions education: BEME Guide No. 30

Mim Fatmi, Lisa Hartling, Tracey Hillier, Sandra Campbell & Anna E. Oswald

To cite this article: Mim Fatmi, Lisa Hartling, Tracey Hillier, Sandra Campbell & Anna E. Oswald (2013) The effectiveness of team-based learning on learning outcomes in health professions education: BEME Guide No. 30, Medical Teacher, 35:12, e1608-e1624, DOI: 10.3109/0142159X.2013.849802

To link to this article: https://doi.org/10.3109/0142159X.2013.849802

Published online: 18 Nov 2013.

Submit your article to this journal

Article views: 5757

Citing articles: 76 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=imte20

2013; 35: e1608–e1624 MEDICAL TEACHER

WEB PAPER
BEME GUIDE

The effectiveness of team-based learning on learning outcomes in health professions education: BEME Guide No. 30

MIM FATMI, LISA HARTLING, TRACEY HILLIER, SANDRA CAMPBELL & ANNA E. OSWALD
University of Alberta, Edmonton, Canada

Abstract

Background: Team-Based Learning (TBL) is a student-centred active learning method, requiring less faculty time than other active learning methods. While TBL may have pedagogical value, individual studies present inconsistent findings. The aim of this systematic review was to assess the effectiveness of TBL on improving learning outcomes in health professions education.

Methods: A peer-reviewed systematic review protocol was registered with the Best Evidence in Medical Education (BEME) organization. After comprehensive literature searching, title and full-text review were completed by two independent reviewers. Included studies assessed TBL and a valid comparator in health professions. Included studies were assessed for methodological quality by two independent reviewers. Studies were categorised by outcomes using the Kirkpatrick framework.

Results: Of 330 screened titles, 14 were included. Seven studies reported significant increase in knowledge scores for the TBL group, four reported no difference and three showed improvement but did not comment on statistical significance. Only one study reported significant improvement in learner reaction for the TBL group while another study reported a significant difference favouring the comparator.

Conclusions: Despite improvement in knowledge scores, there was mixed learner reaction. This may reflect the increased demands on learners in this student-centred teaching strategy, although further study is needed.

Introduction

With increasing enrolment in health professions programmes, there is a growing interest in active learning strategies due to the belief that active learning results in enhanced knowledge retention and skills application (Forsetlund et al. 2009). While traditional lecture-based learning can be successful in teaching students to recall information, health professions students must also be able to think critically and apply their knowledge in novel situations (Parmelee & Michaelsen 2010).

Several active learning methods have been investigated in the past decades. Problem-based learning (PBL) was one of the first to be implemented in health professions education, starting with McMaster University in 1969 and surviving many curricular revisions since then (Koles et al. 2010). Other active learning strategies currently being used include case-based group discussion, workshops and audience response systems. While these methods have proven to be effective in increasing student engagement, they can also be extremely resource-intensive. With medical education facing a crisis in faculty time for teaching, alternative methods that do not require high faculty to student ratios are being sought.

Team-based learning (TBL) attempts to balance the issues of active learning and faculty teaching time (Searle et al. 2003). By breaking up a large lecture hall of 100 or more students into small groups, TBL gets students actively learning while only requiring one faculty to facilitate (Koles et al. 2005). TBL was developed by Dr. Larry Michaelsen in a business curriculum in the 1970s. Michaelsen assigned the students

Practice points

- Team-based learning appears to improve knowledge scores but yields mixed positive and negative learning reaction.
- Curriculum planners who do implement TBL are advised to take precautions to mitigate potentially negative learner reactions to this teaching strategy. The authors hypothesise that this may be due to increased student workload and shift in culture towards peer assessment and accountability; however, further research is needed to clarify the reasons for low learner satisfaction.
- TBL appears to be promising over various settings and populations within health professions education in this small group of studies.
- More robust, controlled primary studies, with thorough reporting and with inclusion of higher lever learning outcomes such as application skills and behaviours, would be helpful to establish a stronger evidence base for curriculum planners considering implementation of TBL.

Correspondence: Anna Oswald, 562 Heritage Medical Research Centre, University of Alberta, Edmonton, AB T6G 2S2, Canada. Tel: 780 407 8711; fax: 780 407 6055; email: oswald@ualberta.ca

ISSN 0142–159X print/ISSN 1466–187X online/13/121608–17 © 2013 Informa UK Ltd.
DOI: 10.3109/0142159X.2013.849802

Copyright Informa
Not for Sale or Commercial
Unauthorized use prohibited
Authorized user may download, display, view and print single copy for personal use.

into teams of 5–7 and informed them that he would not lecture but instead oversee their discussions, ensure that they covered the course content and create challenging problems to which they would apply their knowledge (Michaelsen et al. 2008). Michaelsen refined the process of TBL over the years to characterise it with three main phases: (1) advanced preparation by the students, (2) individual and group readiness assurance and (3) application, including discussion and analysis with the entire class (Koles et al. 2010). The permanence of teams, immediate feedback and a meaningful peer evaluation process are additional cornerstones of this learner-centred educational strategy (Michaelsen & Parmelee, personal communication, 2011). After consulting with Dr. Larry Michaelsen, founder of TBL, and Dr. Dean Parmelee, a second key expert in the field, we have used this model of TBL in our inclusion criteria because we feel this definition captures the essence of TBL most holistically. We acknowledge that other TBL models exist (e.g. hybrid models), but excluded studies of these models on the basis that their outcomes could not be a true evaluation of TBL if only certain elements of TBL were used. As this systematic review protocol was registered in 2011, subsequent published definitions were not included in this review.

The first reported implementation of TBL in health professions education was at the Baylor College of Medicine in 2001 (Haidet et al. 2002). Within one year, 10 medical institutions in the United States piloted TBL (Searle et al. 2003). Currently, TBL is being used at schools of medicine, nursing, dentistry, pharmacy, residency programs and continuing medical education in various jurisdictions, including Japan, Korea, Singapore and the Middle East (Parmelee & Michaelsen 2010).

While the primary literature on TBL describes some advantages in particular settings, no systematic review has been done assessing the effectiveness of TBL on learning outcomes in health professions education. Some non-systematic reviews explore how to best implement TBL, personal experiences with TBL and TBL combined with e-learning (Michaelsen et al. 2008; Davidson 2009; Parmelee & Michaelsen 2010). However, these reviews address one aspect of a TBL experience rather than providing a holistic view of whether TBL serves to improve learning outcomes.

A more thorough review has been done on the effectiveness of several active learning strategies; however, this review was specific to engineering education (Prince 2004). The review included studies on collaborative learning, cooperative learning and TBL. The author concluded that support can be found for all forms of active learning and provided some evidence to promote team-based, collaborative learning environments (Prince 2004). Health professions curriculum planners do not have a resource that evaluates whether or not health professionals' learning outcomes improve with TBL. There is a demand for such a resource due to the increasing use of TBL in health professions education. Some of the specific features of clinical teaching are unique to health professions education. For example, health professions students require preparation for intense workplace-based learning which is often in a high stakes setting that requires attention to multiple domains of competency such as team work and collaboration (Amin & Eng 2009). In addition, elements like clinical reasoning, patient-based teaching and tripartite interaction between students, teachers and patients set health professions education apart from other educational disciplines (Amin & Eng 2009).

Despite the increasingly wide use of TBL in health professions education, no systematic review has been done to rigorously assess the effectiveness of this teaching strategy. We conducted a systematic review to examine the totality of evidence on the effectiveness of TBL in the distinct setting of health professions education. A synthesis of the evidence can provide direction for those in positions of curriculum design and resource allocation.

Research question

The research question for this systematic review was: ''Is TBL effective in improving learning outcomes in health professions education?'' For the purpose of this review, TBL was defined as above to limit heterogeneity between TBL models included in the review.

Methods

Search strategy

A comprehensive search strategy was developed by a health science librarian (SC) in consultation with the co-authors. We identified relevant studies from the electronic databases listed in Table 1. Two search strategies were used depending on whether the database in question was health related or not. This was done to ensure the identification of all relevant studies. The specific terms and search strategies can be found for health-related databases in Table 2 and general databases in Table 3. In addition, the reference lists of all included studies

Table 1. Included online databases.

Health-related databases	General databases
Medline (1950–present)	Physical Education Abstracts
EMBASE (1980–present)	SCOPUS (1823–present)
PubMed (1950–present)	Web of Science (1956–present)
CINAHL (1937–present)	ERIC (1966–present)
Cochrane Library (various dates–present)	OpenSigle (various years–present)
	Proquest Dissertations and Theses (content dates vary - present)

Databases – note that all searches were limited to 1970 to current (July, 2011).

Table 2. Search terms and strategy for health-related databases.

Health trainee education methods	and ''TBL''
exp Education/ or exp Educational Technology/ or ''teaching method*''.mp. or curriculum.mp. or ''instructional method*''.mp.	''team based learn*'' or (tbl and team*)

Limits: English language, human, 1970–present

Table 3. Search terms and strategy for general databases.

Team-based learning	And ''Health professions''
''team based learn*''	medic*
or (TBL and team*)	or nurs*
DE ''TEAM learning approach in education"	or ''physical therap*"
	or physician*
	or health
	or dentist*
	or pharmac*
	or ''occupational therap*''
	or doctor*
	or dietitician*
	or psychologist*
	or psychiatr*
	or clinic*
	or clerkship*
	or veterinary*

Limits: English language, human, 1970–present.

were hand searched, as were those of relevant reviews that were identified during the title screening procedure described below. We also hand-searched conference proceedings for the Association of American Medical Colleges, the Association of Medical Education in Europe, the Canadian Conference of Medical Education and the Team-based Learning Collaborative from 2009 to 2011. A separate cited reference search was conducted using Web of Science and SCOPUS for each included study to identify papers where it had been cited. The primary authors of all included studies were contacted through email to determine if they knew of any unpublished, recently published or ongoing studies relevant to the review.

Screening and selection of studies

The titles and abstracts generated from the electronic database searches were collated in a RefWorks reference management database. They were then screened by two reviewers (AO and MF) to exclude those that obviously did not meet the inclusion criteria or address the question under study (Table 4). For any abstracts that did not provide enough information on the study to determine if they meet our inclusion criteria, the primary authors of the study were contacted twice for further information, after which the study was excluded if we did not receive clarification. The full texts of the remaining studies were retrieved and a pre-approved inclusion form was applied independently by two reviewers (AO and MF) to each to identify relevant studies. Disagreements were resolved through discussion, or with the aid of a third reviewer (LH) as required.

Assessment of methodological quality

The methodological quality of included studies was evaluated independently by two reviewers (TH and MF) using the Cochrane Risk of Bias tool for controlled trials (Higgins & Green 2009) and the Newcastle–Ottawa Scale (NOS) for cohort studies (Wells et al. 2000). These tools are commonly used in systematic reviews (Wells et al. 2000; Hartling et al. 2012) to assess the methodological quality of primary studies and are recommended by The Cochrane Collaboration (Higgins & Green 2009). The Cochrane Risk of Bias tool assesses six general domains that have empirical evidence demonstrating their association with biased estimates of effect: sequence generation, allocation concealment, blinding (participants/personnel, outcome assessment), incomplete outcome data, selective outcome reporting and ''other sources of bias'' (e.g. baseline imbalances between groups, design specific items for cross-over or cluster trials). Domains are rated as low, unclear or high risk of bias. Overall risk of bias can be assessed for each study: studies are considered low risk of bias if all individual domains are assessed as low, high risk of bias if one domain is assessed as high and unclear otherwise. The Newcastle–Ottawa Scale for cohort studies assesses the selection of participants, comparability of study groups and ascertainment of the outcome of interest. A rating system is used to indicate the overall quality of a study with a maximum assessment of nine points (Wells et al. 2000; Hartling et al. 2012). Discrepancies were resolved through consensus or with the aid of a third reviewer (LH) as required.

Data extraction

Data were extracted and entered into an electronic data extraction form. The form was developed and piloted in a systematic review performed by the authors (Hartling et al. 2010), and further revised and tailored to the current review. One reviewer extracted data (MF). To ensure accuracy and consistency a 20% sample of the articles was randomly selected for extraction by a second reviewer (TH). The data extracted by the two reviewers was compared and no significant discrepancies or errors were detected.

Analysis

Our inclusion criteria allowed for a mixture of qualitative and quantitative studies. The data were synthesized with guidance from methods described by Ogawa and Malen (1991) who have developed a method for synthesis based on the exploratory case study method. Thus, evidence was iteratively synthesised by grouping studies by various constructs including by the population involved and nature of the comparison group, and summarised according to the reported outcomes which were grouped by Kirkpatrick level. The results of this synthesis were then described, categorised and summarised in Table 5. Evidence tables detailing study characteristics (including population, intervention, comparison, outcomes, design and any modifications to classic TBL) (Table 6), results and authors' conclusions are provided (Table 7). Statistical meta-analysis was not performed because of substantial heterogeneity across the intervention type, comparator, study designs and insufficient reporting of data at the study level. Conclusions about the effectiveness of TBL were drawn based on review of results from studies reporting similar outcomes.

Results

Figure 1 presents a flow diagram of the study selection process. Three hundred and thirty studies were identified, and 14 were included (Table 6). Among the included studies were one RCT (Koles et al. 2005), two NRCTs (Thomas & Bowen

Table 4. Inclusion and exclusion criteria applied to potentially relevant studies to determine suitability for systematic review purposes.

	Inclusion criteria	Exclusion criteria
Population	Medical Students Residents Physicians Nursing Students/Nurses Pharmacy Students/Pharmacists Dental Students/Dentists Veterinary Trainees/Veterinarians Dietician Trainees/Dieticians Clinical Psychology Trainees/Clinical Psychologists Other Allied Health Professionals Interdisciplinary health professions teams	Non-Health Professions Trainees
Intervention	Team-based learning in conjunction with: Lectures Workshops Small group learning sessions Clinical teaching Other structured teaching sessions	Shadowing Mentoring Practice audits Feedback alone
Comparator	Any teaching method described under the inclusion criteria for Intervention section without TBL. Any ''standard curriculum'' without TBL	
Outcome (Categorised by the modified Kirkpatrick's 1967 model of hierarchical outcomes (Kirkpatrick & Kirkpatrick 2006))	Change in patients' health Change in behaviour Inclusion of skill in clinical practice Change in skills OSCE scores Observed assessment scores Change in knowledge Written exam scores Change in attitudes/perceptions Confidence self ratings Comfort self ratings Learner Reaction Satisfaction with teaching method Satisfaction with instructor	
Study type	Comparative studies which provide primary data for any of the outcomes listed above, including the following designs: Randomised controlled trials Non-randomised control trials Controlled before and after studies Interrupted time series Other robust comparative studies (e.g. cohort studies) Qualitative comparative studies English language (Morrison et al. 2009)	Studies reporting on needs assessments for TBL Studies reporting the prevalence of TBL Opinion Papers Studies without a comparator group Studies with uncontrolled before and after design Articles not in the English language

2011; Willett et al. 2011), two prospective cohort studies (Torralba et al. 2009; Wiener et al. 2009), one retrospective cohort study (Koles et al. 2010), one concurrent cohort study (Zingone et al. 2010) and seven non-concurrent cohorts (Nieder et al. 2005; Levine et al. 2004; Letassy et al. 2008; Pileggi & O'Neill 2008; Mennenga 2010; Simaan et al. 2010; Zgheib et al. 2010). Of the 14 included studies, 12 took place in the United States, 1 in Lebanon (Zgheib et al. 2010) and 1 in Austria (Wiener et al. 2009). Thirteen of the studies concerned undergraduate education, including eight in medicine, three in pharmacy (Letassy et al. 2008; Conway et al. 2010; Zgheib et al. 2010), one in dentistry (Pileggi & O'Neill 2008) and one in nursing (Mennenga 2010). The only study that did not include undergraduate learners assessed internal medicine residents in their first, second and third postgraduate years (Torralba et al. 2009). All 14 studies assessed knowledge as a learning outcome, and 7 studies also assessed learner reaction (Koles et al. 2005; Levine et al. 2004; Letassy et al. 2008; Conway et al. 2010; Zingone et al. 2010; Thomas & Bowen 2011; Willett et al. 2011). None of the studies evaluated Kirkpatrick outcomes such as skills or improvements for patients. In total, the studies included over 3535 participants (exact numbers are not known as three studies did not report the number of control group participants) of which 1869 students received TBL sessions.

Due to the nature of TBL, all studies assessed knowledge scores for the TBL groups during the teaching session in the

Table 5. Summary of findings.

Outcome	Intervention	Comparator	Findings: Any significant difference			Study design and number of participants enrolled
			No statement	$p > 0.05$	$p < 0.05$	
Knowledge	TBL	CBGD		No difference		RCT ($n = 83$)
					Favours TBL	NCC ($n =$ unclear)
		SGL			Favours TBL	NRCT ($n = 112$)
					Favours TBL	NRCT ($n = 167$)
		Mixed Active Learning			Favours TBL	CC ($n = 64$)
		Independent Study	Favours TBL			PC ($n = 1417$)
		Traditional Lecture		No difference		NCC ($n =$ unclear)
					Favours TBL	RC ($n = 186$)
			Favours TBL			NCC ($n = 280$*)
					Favours TBL	NCC ($n = 306$)
				No difference		NCC ($n = 143$)
					Favours TBL	NCC ($n = 371$)
			Favours TBL			NCC ($n =$ unclear)
				No difference		PC ($n = 121$)
Reaction	TBL	CBGD		No difference		RCT ($n = 83$)
		SGL		No difference		NRCT ($n = 112$)
					Favours SGL	NRCT ($n = 167$)
		Mixed Active Learning		No difference		CC ($n = 64$)
		Traditional Lecture	Favours TBL			NCC ($n =$ unclear)
			Favours lecture			NCC ($n = 280$*)
					Favours TBL	NCC ($n = 306$)

RCT = randomized controlled trial; NRCT = non-randomized controlled trial; NCC = non-concurrent cohort; CC = concurrent cohort; PC = prospective cohort; RC = retrospective cohort.
*The exact number of participants enrolled in the study was not reported.

form of IRATs and/or GRATS. However, because most alternative teaching methods do not require teachers to assess students' knowledge during the actual teaching session and therefore do not report this short-term knowledge score, there were no comparative data between TBL and non-TBL groups for this short-term knowledge variable. All included studies provided knowledge scores comparing groups for the course or the semester in which TBL was implemented and so these are the variables analysed in this review.

Methodological quality and risk of bias of included studies

Quality assessment reveals various methodological shortcomings and these are reported by study in Tables 8 and 9. The three randomised and non-randomised controlled trials were assessed for quality using the Cochrane Risk of Bias tool. Allocation appeared to be appropriately randomised in the RCT (Koles et al. 2005), while the two NRCTs allocated students to groups based on last name (Thomas & Bowen 2011; Willett et al. 2011). None of the trials attempted to blind the students to their allocation or to the study hypothesis. None of the trials seemed to be at risk of selective outcome reporting or other sources of bias.

For 8 of the 11 cohort studies, the learners were truly representative of the average health-care professions student, and both exposed and non-exposed cohorts were drawn from the same community (Nieder et al. 2005; Levine et al. 2004; Letassy et al. 2008; Pileggi & O'Neill 2008; Conway et al. 2010; Mennenga 2010; Zgheib et al. 2010; Zingone et al. 2010). One study lost five failing students, leading the reviewers to deem the remaining cohort somewhat, but not truly, representative of the average health professions class. Two cohort studies assessed TBL within a group of self-selected volunteers (Torralba et al. 2009; Wiener et al. 2009). The majority of studies did not attempt to control for potential confounders between the exposed and non-exposed cohorts with regard to learning aptitudes, histories, etc.; however, one study did analyse scores after taking pre-intervention GPAs into account while another looked at age, health-care experience, previous GPAs and other measures (Mennenga 2010; Zingone et al. 2010). The results observed among the studies that provided adjusted estimates were inconsistent; therefore, it is uncertain whether unadjusted results from the remaining studies may have over- or underestimated intervention effects.

All studies had a clear definition of the outcome being assessed and collected the outcome data via record linkage for knowledge scores and via student self-reporting for learner reaction. Four studies did not report on completeness of follow-up (Letassy et al. 2008; Pileggi & O'Neill 2008; Torralba et al. 2009; Conway et al. 2010). One study appeared to have a loss to follow-up rate of greater than 10% with an unclear explanation of learners lost; however, this was a very large-scale study that still reported a large number of participants despite the loss to follow-up (Wiener et al. 2009).

Characteristics of included studies

Table 5 provides a summary of the interventions, comparators, outcomes measured and main findings of all included studies. Best efforts to limit heterogeneity between the models of TBL were employed. However, for full transparency a column of

Table 6. Study characteristics.

Citation	Institution	Design	Population (*N*)	Research question/ purpose	Intervention (*N*)	Comparator (*N*)	Primary outcomes	Observed modifications to classic TBL
(Koles et al. 2005)	Wright State University School of Medicine	RCT	Undergraduate medical (83)	To evaluate the impact of TBL on the academic performance of Year 2 medical students at Wright State University by comparing this active learning strategy against a traditional method of CBGD	TBL incorporated into eight pathology modules (80)	CBGD (80)	Knowledge: final exam scores Learner reaction	Unclear if peer assessment performed
(Thomas & Bowen 2011)	Johns Hopkins University School of Medicine	NRCT	Undergraduate medical (112)	To compare the impact of TBL versus usual small group lecture (SGL) on knowledge objectives in a clinical clerkship	TBL format used in six ambulatory medicine modules (112)	Small group learning (112)	Knowledge: exam scores Learner reaction	None observed
(Willett et al. 2011)	University of Medicine and Dentistry of New Jersey	NRCT	Undergraduate medical (167)	To determine if team-based learning could be substituted for small group learning in case sessions without compromising test performance or satisfaction	Endocrinology and rheumatology module offered in TBL format (84)	Small group learning (83)	Knowledge: exam scores Learner reaction	Lectures still used to deliver content; RATs and peer assessment did not contribute to students' grade
(Conway et al. 2010)	University of Oklahoma College of Pharmacy, Oklahoma City and Tulsa campuses	Non-concurrent cohort	Undergraduate pharmacy (*N* = unclear)	To integrate components of team-based learning (TBL) into a cardiovascular module to increase students' responsibility for their own learning and actively engage students across two campuses in patient cases	Lecture-based course replaced with self-directed learning (SDL) assignments and case discussions were replaced with TBL in six sessions (140)	Traditional lecture format (*N* = unclear)	Knowledge: number of assigned course grades Learner reaction	No GRATs; lectures still used to deliver content
(Koles et al. 2010)	Wright State University School of Medicine	Retrospective cohort	Undergraduate medical (186)	To determine whether student performance on examinations is affected by participation in TBL and whether TBL benefits lower- or higher-performing students	TBL incorporated into 10 courses (178)	Traditional lecture format (primarily), laboratory, clinical case discussions, and independent study module (178)	Knowledge: final exam scores	Mixed curriculum includes TBL, lectures, labs and clinical sessions, to deliver content

(continued)

Table 6. Continued.

Citation	Institution	Design	Population (*N*)	Research question/ purpose	Intervention (*N*)	Comparator (*N*)	Primary outcomes	Observed modifications to classic TBL
(Letassy et al. 2008)	University of Oklahoma College of Pharmacy, Oklahoma State University Center for Health Sciences	Non-concurrent cohort	Undergraduate pharmacy (280)	To implement a team-based learning (TBL) format in an endocrine module to promote students' active learning in a course delivered to two campuses	Endocrinology module offered in 13 TBL sessions at two campuses (140)	Traditional lecture format (140)	Knowledge: unit exam scores Learner reaction	None observed
(Levine et al. 2004)	University of Texas Medical Branch	Non-concurrent cohort	Undergraduate medical (306)	To compare students' performance and educational experiences before and after curricular change	8 of 16 lecture sessions were replaced with TBL activities; only 4 of 8 sessions implemented IRATs/ GRATs (133)	Traditional lectures (primarily) and case conferences (131)	Knowledge: NBME psychiatry shelf exam scores Learner reaction	TBL slowly phased in; lectures still used to deliver content
(Mennenga 2010)	University of Nevada, Las Vegas	Non-concurrent cohort	Undergraduate nursing (143)	To examine differences in student engagement between baccalaureate nursing students taught using team-based learning and those taught using traditional lecture, and how levels of engagement affected exam scores	Lecture-based community health nursing course replaced with TBL (69)	Traditional lecture (74)	Knowledge: exam scores	None observed
(Nieder et al. 2005)	Wright State University School of Medicine	Non-concurrent cohort	Undergraduate medical (371)	To determine if students would be better prepared and find TBL more effective than previous small group sessions	12 TBL sessions incorporated into a human anatomy and embryology course (95)	Traditional lecture (276)	Knowledge: number of exam failures	Lectures also used to deliver some content
(Pileggi & O'Neill 2008)	University of Texas Health Science Center	Non-concurrent cohort	Undergraduate dental (*N* = unclear)	To facilitate student learning and performance in a sophomore preclinical endodontic course and improve student ability to diagnose diseases utilizing TBL combined with an audience response system	Lecture-based endodontics course replaced with TBL (64)	Traditional lecture (*N* = unclear)	Knowledge: final exam scores	Group size larger (10–11); unclear whether RATs were weighted in students' final grade; GRATs were ''parallel'' but not identical to IRATS; description of application exercise unclear; peer assessment unclear

(Torralba et al. 2009)	University of Southern California School of Medicine	Prospective cohort	Residents (121)	To study the effectiveness and outcomes of using TBL in Rheumatology core curriculum learning sessions (CCLS) for Internal Medicine (IM) residents over an initial period of 12 months	Seven rheumatology CCLS offered in TBL format (68)	Traditional lecture (53)	Knowledge: in-training exam scores	Value of Teams Survey was conducted in place of peer assessment
(Wiener et al. 2009)	Medical University of Vienna	Prospective cohort	Undergraduate medical (1417)	To examine the impact of team-based learning (TBL) on educational outcomes in the first year of the curriculum of the Medical University of Vienna	Optional intensive review course for final exam offered in TBL format (588)	Independent study (829)	Knowledge: final exam scores	The course was an optional intensive review over a 3-day period; participants were self-selected; RATs were not assigned weight to final grade; peer assessment unclear
(Zgheib et al. 2010)	American University of Beirut Faculty of Medicine	Non-concurrent cohort	Undergraduate medical (N = unclear)	To examine the effect of teaching pharmacology using a TBL approach on second year medical students' satisfaction and performance	TBL was incorporated into two CBGD sessions of pharmacology (N = unclear)	CBGD (N = unclear)	Knowledge: exam scores	TBL implemented in only two class sessions (short duration); student-selected groups; lectures also used to deliver content; RATs were not assigned weight to final grade; peer assessment not performed
(Zingone et al. 2010)	University of Tennessee Health Science Center College of Pharmacy	Concurrent cohort	Undergraduate pharmacy (64)	To assess students' performance and perceptions of team-based and mixed active-learning methods in two ambulatory care elective courses	Ambulatory care elective course offered in TBL format (37)	Mixed active learning methods (27)	Knowledge: final grades Learner reaction	None observed

Table 7. Main findings of the review.

Citation	Design	Outcomes	Summary of results	Author's conclusions	Stated limitations
(Koles et al. 2005)	RCT	Knowledge: final exam scores Learner reaction	No significant differences in exam scores were found between the TBL session exams and the CBGD exams. For the immune, neoplastic, cardiovascular and parathyroid modules, the TBL group scored 80.7% and the CBGD scored 81.2% for modules INCP ($p = 0.79$). For the genetic, muscle, breast and liver disease modules, the TBL group scored 85.9% and the CBGD group scored 87.6% ($p = 0.36$). However, in the lowest academic quartile, the authors report a significantly reduced deterioration of long-term knowledge retention after experiencing a TBL session than a CBGD (−1.5% vs. −9.3%, $p = 0.035$). No significant difference was found between learner reaction scores ($p = 0.143$).	''TBL and CBGD are equally effective active learning strategies when employed in a systems-based preclinical pathology curriculum, but students with lower academic performance may benefit more from TBL than CBGD.''	TBL and CBGD are fundamentally different teaching methods.
(Thomas & Bowen 2011)	NRCT	Knowledge: exam scores Learner reaction	A significant difference was observed in exam scores between the TBL and SGL cohorts in five of six modules favouring TBL. The average TBL score for the first three modules was significantly higher than the SGL score (77.1% vs. 73.2%, $p = 0.05$) as well as for the second three modules (78.4% vs. 66.8%, $p < 0.001$). The authors reported an insignificant difference between learner reaction to the usefulness of the modules taught via TBL or SGL (4.35 vs. 4.29); no p value was given.	''This study adds to the growing body of evidence suggesting effectiveness of the TBL method in achievement of knowledge objectives for the health professional student.''	This study took place in one medical school with one faculty facilitator; expertise bias could have been introduced; the clerkship was 4 weeks, which did not allow for the optimal 40-hour investment to develop teams.
(Willett et al. 2011)	NRCT	Knowledge: exam scores Learner reaction	There was a significant difference between the exams score of TBL and SGL groups favouring TBL (81.7% vs. 79.7%, $p = 0.04$). However, the groups also performed significantly differently on modules in which TBL was not implemented (79.5% vs. 77.1%, $p = 0.01$). After adjusting for performance on the non-TBL modules, the authors reported no significant difference. Students significantly preferred SGL over TBL (3.74/5 vs. 2.45/5, $p < 0.001$).	''We found that a TBL approach to case-based learning was feasible and resulted in similar examination performance when compared to a concurrent SGL approach.''	The study was implemented in one course at one institution; TBL was not the only teaching strategy used; allocation to groups was not randomised; only 43% of students responded to the surveys.

(Conway et al. 2010)	Non-concurrent cohort	Knowledge: number of assigned course grades Learner reaction	No statistically significant differences were found between student performance before and after TBL implementation; however, the authors report a non-significant trend towards increasing average exam scores after implementation. The number of students receiving non-passing grades (D and F) decreased after TBL implementation (2.7% vs. 0%) and the number of students receiving B's increased (58.9% vs. 63.0). As well, the number of students receiving A's decreased after TBL (19.6% vs. 17.8%), though no p values were given. Average learner satisfaction scores are reported to have increased after TBL was implemented (4.4 vs. 4.2); no p value was given.	''Compared to pre-TBL cohort, student performance was maintained and satisfaction was increased.''	NS
(Koles et al. 2010)	Retrospective cohort	Knowledge: final exam scores	TBL-related questions were answered correctly an average of 5.9% more often than TBL-unrelated questions. 83.6% of TBL-related questions were answered correctly while 77.9% of TBL-unrelated questions were answered correctly ($p < 0.001$). Within the lowest academic quartile, TBL-related questions were answered correctly more often than in the highest academic quartile (7.9% vs. 3.8%, $p = 0.001$).	''Overall, students in two consecutive second-year classes demonstrated significantly higher performance on pathology based questions 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.'' 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.	The study was limited to pathology; did not compare TBL with another active learning strategy; and only short-term knowledge was tested.
(Letassy et al. 2008)	Non-concurrent cohort	Knowledge: unit exam scores Learner reaction	Though no p value or statistical difference was reported, the TBL cohort scored higher than the lecture cohort (86% vs. 81%). Learner reaction scores were lower for TBL (3.5/5 vs. 4.6/5); no p value was given.	''TBL effectively covers course content without sacrificing grades. The repetitive nature of the TBL process appears to provide the most benefit to academically weak students due to the increased structure and accountability for learning.''	NS
(Levine et al. 2004)	Non-concurrent cohort	Knowledge: NBME psychiatry shelf exam scores Learner reaction	There was a significant increase in NBME psychiatry subject exam scores between the TBL cohort and lecture cohort (72.9% vs. 69.6%, $p < 0.05$). There was also a significant difference in learner reaction favouring TBL (3.68/5 vs. 3.21/5, $p < 0.001$).	''The results suggest that team learning activities could be an important addition to traditional clerkship didactic programs in medical education. Attainment of core knowledge, as determined by NBME subject test scores, was significantly higher than in the previous curriculum. Team learning activities were perceived to be significantly more effective and enjoyable.''	Gathering controls from previous academic year is not ideal; attitudes could have been influenced by different faculty.
(Mennenga 2010)	Non-concurrent cohort	Knowledge: exams scores	No significant difference was found on exam scores for four exams between the TBL group and lecture group (41.32/50, 42.26/50, 31.39/40, 49.72/60 vs. 40.99/50, 39.07/50, 34.85/40, 49.66/60, $p = 0.923$).	While some findings did not suggest team-based learning to be better than traditional lecture, the findings regarding examination scores do suggest that team-based learning is at minimum equally as effective as traditional lecture.	The population size of 143 is relatively small; the author had limited experience with TBL; differences could have existed between the control and experimental groups; the validity of instrument assessing the effectiveness of TBL is relatively new.

(continued)

Table 7. Continued.

Citation	Design	Outcomes	Summary of results	Author's conclusions	Stated limitations
(Nieder et al. 2005)	Non-concurrent cohort	Knowledge: number of exam failures	There was a significant decrease in the number of exam failures after the implementation of TBL compared to the lecture group (1 vs. 6, $p < 0.001$). The authors reported no significant difference between the average exam scores before and after TBL implementation ($p = 0.344$).	Although the class's grade average was similar to previous years' averages, we believe that the team-based learning did help the few academically weaker students to succeed.	NS
(Pileggi & O'Neill 2008)	Non-concurrent cohort	Knowledge: final exam scores	The final exam scores increased after the implementation of TBL compared to the didactic lecture group one and two years prior (75% vs. 69% and 56%, respectively); no p value was given.	''The course faculty believe that this methodology improved students' didactic knowledge as well as their diagnostic skills when in clinic. We believe that using TBL enhanced our students' skills on differential diagnosis and hopefully prepared them for success in their daily clinical education experience.''	NS
(Torralba et al. 2009)	Prospective cohort	Knowledge: in-training exam scores	The average exam scores for residents in the TBL group were slightly higher than the didactic lecture group, but did not reach significance (54.4% vs. 50.7%, $p = 1.10$). The authors also analysed the differences between the groups by year of residency and found a significant difference at the PGY1 level ($p = 0.03$), though no significant differences were found at the PGY2 or PGY3 levels.	''Internal medicine residents are receptive to the use of TBL for CCLS, and have favourable perceptions of collaborative learning. Additionally, results provide information on potential impact of TBL on available assessments such as the ITE.''	NS
(Wiener et al. 2009)	Prospective cohort	Knowledge: final exam scores	Students who elected to take the TBL review course scored 18.3% higher than the students who did not (137/230 vs. 116/230); no p value was given. The students who took the TBL course also had a 16.5% higher score in the remaining five non-TBL blocks of the year when compared to the students who did not take the elective.	''Providing TBL as an elective fosters a positive selection of the best students of the year making them even more successful on the final exam.''	The students self-selected to take the elective TBL course.
(Zgheib et al. 2010)	Non-concurrent cohort	Knowledge: exam scores	Student performance on exams improved significantly in the module in which TBL was implemented compared to the didactic lecture groups one and two years prior (75% vs. 58% and 47%, respectively, $p = 0.03$).	''Performance of the students on the summative tests was improved in students who took TBL relative to those in previous classes, who were taught in a more traditional manner.''	The method of TBL used was modified; TBL was only implemented in two sessions; and the study did not compare with PBL or CBGD.
(Zingone et al. 2010)	Concurrent cohort	Knowledge: final grades Learner reaction	Students in the TBL group scored a significantly higher average GPA than the mixed active learning (MAL) group (3.7/4 vs. 3.3/4, $p < 0.001$). After adjusting for GPA prior to course entry, the authors determined the TBL group had earned an average 0.33 GPA points higher than the MAL group. Learner reaction was similar in both groups, though improved confidence was rated slightly, but insignificantly, lower in the TBL group (3.4/4 vs. 3.5/4, $p = 0.900$).	''Student performance, measured by quality points, was significantly higher in the team-based learning course format when adjusted for incoming GPAs. Students were satisfied with both learning models.''	Some students perform more favourably on different types of assessments; the content covered in each course varied slightly.

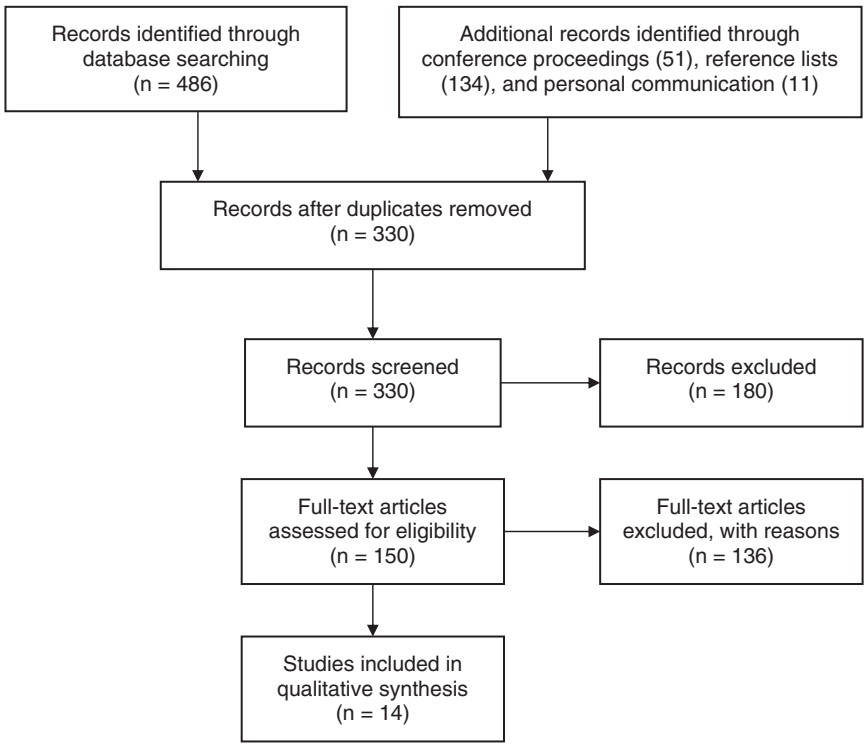

Figure 1. Flow diagram of included studies.

any minor discrepancies between our definition of TBL and the model implemented in the study are described in Table 6. Tables 6 and 7 describe the results of all included studies. The following narrative provides a summary of the findings grouped by outcome.

Knowledge

All 14 studies evaluated knowledge-based learning outcomes in a TBL group compared with a non-TBL group in a health education curriculum, assessing a total of at least 3535 participants; 3 studies were unclear about the number of participants in their historical controls. All studies reported knowledge outcomes and these were measured within the semester or course in which TBL was implemented. Seven studies reported a statistically significant increase ($p < 0.05$) in knowledge scores for the TBL group compared with a non-TBL group. However, one NRCT ($n = 167$) reported that after adjusting for pre-intervention knowledge scores between groups, there was no significant difference after TBL implementation (Willett et al. 2011). Among the other studies reporting a significant difference were one NRCT ($n = 112$) (Thomas & Bowen 2011), one retrospective cohort ($n = 186$) (Koles et al. 2010), one concurrent cohort ($n = 64$) (Zingone et al. 2010) and three non-concurrent cohorts ($n = 306$, $n = 371$, $n =$ unclear) (Nieder et al. 2005; Levine et al. 2004; Zgheib et al. 2010). Two studies (one RCT and one prospective cohort) did not report a significant difference between TBL and comparators overall, but found a significant difference in subgroup analyses (Koles et al. 2005; Torralba et al. 2009). Koles et al. ($n = 83$) found significantly improved knowledge retention in students in the lowest academic quartile after experiencing a TBL session compared with CBGD ($p = 0.035$) (Koles et al. 2005). Torralba et al. ($n = 121$) found a significant improvement in resident in-training exam scores in favour of the TBL groups over the lecture groups within Year 1 residents ($p = 0.03$), but not Year 2 or 3 residents (Torralba et al. 2009).

Four studies reported no statistically significant difference between knowledge scores of the TBL and non-TBL group ($p > 0.05$). Of this group are the two studies that did find differences in subgroup analyses mentioned above (Koles et al. 2005; Torralba et al. 2009), and two non-concurrent cohort studies that reported no difference at all ($n =$ unclear, $n = 143$) (Conway et al. 2010; Mennenga 2010). These studies included undergraduate studies in pharmacy (Conway et al. 2010) and nursing (Mennenga 2010).

Three studies did not report on significance testing and did not provide sufficient data for calculations; however, in all three studies the authors concluded that TBL was more effective based on some differences observed between groups (see Table 7) (Letassy et al. 2008; Pileggi & O'Neill 2008; Wiener et al. 2009). Two of these studies were non-concurrent cohort studies ($n = 280$, $n =$ unclear) (Letassy et al. 2008; Pileggi & O'Neill 2008) and one was a prospective cohort study ($n = 1417$) (Wiener et al. 2009). The subjects of these studies were undergraduate studies in pharmacy (Letassy et al. 2008), dentistry (Pileggi & O'Neill 2008) and medicine (Wiener et al. 2009).

Comparison group

The choice of comparator did not appear to influence whether knowledge scores favoured TBL or any one alternate teaching method. Eight studies compared TBL to traditional didactic lectures, three of which reported significant improvements in knowledge scores for the TBL group (Nieder et al. 2005;

Table 8. Methodological quality of randomised controlled trials (RCTs) using the Cochrane Collaboration's tool for assessing risk of bias.

Author (Year)	Sequence generation	Allocation concealment	Blinding		Incomplete outcome data*	Selective outcome reporting	Other
			Participants*	Outcome assessment*			
Koles et al. (2005)	Low	High	High	High	Low	Low	Low
Thomas & Bowen (2011)	High	High	High	High	Low	Low	Low
Willett et al. (2011)	High	High	High	High	Low	Low	Unclear

*Domains for which assessments are made by outcome were assessed for objective outcomes.

Levine et al. 2004; Koles et al. 2010). Two studies compared TBL to CBGD; one study found statistically significant improvements, while the other found significant improvements only in students in the lowest academic quartile (Zgheib et al. 2010; Koles et al. 2005). Two trials compared TBL to their own form of SGL and both reported statistically significant results favouring TBL (Thomas & Bowen 2011; Willett et al. 2011). The remaining two studies used mixed active learning methods and independent study as comparators, which are less standardised learning strategies. While both reported increases in knowledge scores for the TBL group, the latter study did not report results of significance testing (Wiener et al. 2009; Zingone et al. 2010).

Learner reaction

Seven studies, involving at least 1152 participants (726 of whom received TBL), reported controlled learner reaction scores between a TBL and non-TBL group (one study did not report the number of control group participants). Only one cohort study reported a significant difference ($p < 0.05$) favouring the TBL group (Levine et al. 2004). In one NRCT, students significantly preferred the alternative (SGL) to TBL ($p < 0.001$) (Willett et al. 2011). These two studies demonstrate that students had a positive reaction to TBL despite the more abrupt increase in workload when compared to a traditional lecture; however, when students were comparing to a less structured active learning strategy, such as SGL, they did not react as positively to TBL. Three studies reported non-significant differences, of which one was an RCT ($n = 83$) (Koles et al. 2005), one was an NRCT ($n = 112$) (Thomas & Bowen 2011) and one was a concurrent cohort reporting a trend in learner preference for the alternate teaching strategy ($n = 64$) (Zingone et al. 2010). Two studies did not comment on statistical significance ($n =$ unclear, $n = 280$): one study favoured TBL and the other favoured the lecture comparator (Conway et al. 2010; Letassy et al. 2008).

Studies in which a recurrent TBL curriculum was implemented were not found in the literature search, and therefore the authors were not able to determine if learner reaction scores might improve over time as students become familiar with this new learning strategy.

Discussion

This is the first systematic review that we are aware of that examines the effects of TBL in health professions education. Previous reviews of TBL either did not study health education (Prince 2004) or were not full systematic reviews (Parmelee & Michaelsen 2010). The health professions educational setting represents a truly distinct population with specific needs and resources. This review was rigorous in its inclusion eligibility, especially with regard to what constitutes TBL; we ensured that the studies presented here complied with a definition of TBL that was verified with founders and leaders of the TBL community. By including only controlled studies, this review provides evidence on how TBL compares to traditional teaching strategies.

The purpose of this systematic review was to evaluate the effectiveness of TBL in improving learning outcomes in health professions education. The results show both positive and neutral effects on knowledge scores, while learner reaction towards TBL was mixed. These findings are beneficial to those who seek evidence suggesting the effectiveness of TBL in achieving the same, if not better, knowledge objectives as more traditional methods. However, curriculum planners who do implement TBL are advised to take precautions to mitigate potentially negative learner reactions to this teaching strategy.

Fourteen studies were included in the analysis and half (7) of these studies reported a statistically significant increase ($p < 0.05$) in knowledge scores favouring TBL (Koles et al. 2005; Nieder et al. 2005; Levine et al. 2004; Zgheib et al. 2010; Zingone et al. 2010; Thomas & Bowen 2011; Willett et al. 2011). These seven studies reporting a significant difference ($p < 0.05$) varied in their study designs. Three studies did not comment on comparisons of statistical significance between groups, despite a trend in knowledge outcomes favouring TBL (Letassy et al. 2008; Pileggi & O'Neill 2008; Wiener et al. 2009); it is possible that a greater proportion of included studies would have favoured TBL had these three studies been reported more thoroughly. The four studies reporting no statistically significant difference between the study groups ($p > 0.05$) were robust and scored high on quality assessment (see Tables 8 and 9).

The learner reaction findings after the implementation of TBL require more careful consideration. Of seven studies reporting a controlled comparison of learner reaction, only one reported a statistically significant difference in learner reaction favouring TBL (Levine et al. 2004), while one study favoured the comparator (Willett et al. 2011). However, the Willett et al. results must be interpreted with caution as in this study the traditional lecture continued to be the primary instruction modality and the TBL-associated scores, including

Table 9. Methodological quality of prospective cohort studies (PCS) and retrospective cohort studies (RCS) using Newcastle–Ottawa Quality Assessment Scale.

					Comparability of cohorts (Study controls)					
Author (Year)	Representativeness of exposed cohort	Selection of non-exposed cohort	Ascertainment of exposure	Outcome of interest not present at start of study	Known factors	Additional factor	Assessment of outcome	Follow up long enough for outcomes to occur	Adequacy of cohort follow-up	Total stars (quality rating)
Conway et al. (2010)	Truly representative	Same community as exposed cohort	Secure record	No	No	No	Record linkage	Yes	Subjects lost unlikely to introduce bias	7 (good)
Koles et al. (2010)	Somewhat representative	Same community as exposed cohort	Secure record	No	No	No	Record linkage	Yes	Subjects lost unlikely to introduce bias	6 (fair)
Letassy et al. (2008)	Truly representative	Same community as exposed cohort	Secure record	No	No	No	Record linkage	Yes	No statement	5 (fair)
Levine et al. (2004)	Truly representative	Same community as exposed cohort	Secure record	No	No	No	Record linkage	Yes	Complete follow up	6 (fair)
Mennenga (2010)	Truly representative	Same community as exposed cohort	Secure record	No	Yes	Yes	Record linkage	Yes	Complete follow up	8 (good)
Nieder et al. (2005)	Truly representative	Same community as exposed cohort	Secure record	No	No	No	Record linkage	Yes	Subjects lost unlikely to introduce bias	6 (fair)
Pileggi & O'Neill (2008)	Truly representative	Same community as exposed cohort	Secure record	Yes	No	No	Record linkage	Yes	No statement	6 (fair)
Torralba et al. (2009)	Self-selected group	Same community as exposed cohort	Secure record	No	No	No	Record linkage	Yes	No statement	4 (fair)
Wiener et al. (2009)	Self-selected group	Same community as exposed cohort	Secure record	No	No	No	Record linkage	Yes	Loss to follow up >10%	4 (fair)
Zgheib et al. (2010)	Truly representative	Same community as exposed cohort	Secure record	No	No	No	Record linkage	Yes	Subjects lost unlikely to introduce bias	6 (fair)
Zingone et al. (2010)	Truly representative	Same community as exposed cohort	Secure record	No	No	Yes	Record linkage	Yes	Subjects lost unlikely to introduce bias	7 (good)

the GRAT and peer evaluation, did not contribute to the students' grade; thus, there may have been less student engagement in the TBL process. Of the four studies that reported both a significant increase in knowledge scores for TBL and also reported controlled learner satisfaction data, only one study reported a significant increase in learner reaction (Levine et al. 2004). We hypothesise that these conflicting knowledge and learner satisfaction results may be due to the strain of increased workload associated with TBL (e.g. required advanced readings and preparation) and a change in the professional culture towards peer assessment and accountability that TBL introduces. These hypotheses are supported by some of the authors' (AO and TH) recent experiences of TBL implementation in a preclinical medical student course on musculoskeletal medicine at our home university. However, further studies are required to confirm and shed further light for the reasons of poor learner satisfaction.

The results for knowledge outcomes appeared to be consistent regardless of the type of comparator used in the included studies. The majority of comparison groups were comprised of traditional lectures, while CBGD and SGL were also common alternative teaching methods to TBL. There was a suggestion in the data that learner reactions were more positive in favour of TBL when traditional lectures were used as the comparator rather than less structured small group learning strategies, but further study is needed.

The limitations of this review were minimised with regard to the review design by prospectively establishing thorough parameters for database searches, by having two reviewers screen, reconcile and assess the quality of potentially relevant studies, and by contacting authors for any recently published or unpublished data. The major limitations of this review are due to the methodological quality of the included studies. Most of the included studies were cohort designs, and the majority of them were non-concurrent. Further, the authors of these convenience studies often did not report full statistical information or sufficient data to allow for statistical comparisons. Many studies did not control for differences between groups to prevent confounding. In an attempt to explore and better understand the effects of poor study quality, a thorough quality assessment of included studies was performed and reconciled by two reviewers. Although the more robust trials reported data more thoroughly, it is reassuring that their findings were as mixed as those of the cohort studies, indicating that study design was unlikely to be a major source of bias in this review.

A second potential limitation relates to the constraints of the inclusion criteria for this review. One constraint was restricting inclusion to studies that fully complied with the validated definition of TBL that was accepted in the registered protocol for this review. Overall, this resulted in the exclusion of certain robust studies implementing forms of TBL that did not meet the criteria in the protocol, for example, those that omitted a major component such as readiness assurance testing. However, we felt that this strategy prevented dilution of the results by allowing extreme heterogeneity of interventions.

As with most studies in the field of education, there are inherent limitations with regard to asserting one strategy's effectiveness over another. Prince et al. outline the problems on interpreting the literature of active learning (Prince 2004). Although we have attempted to address Prince et al.'s noted difficulty in defining exactly what is being studied, there are still challenges in measuring what works (Prince 2004). The latter difficulty is particularly relevant in this review, as the outcomes measured in the included studies are limited to learner reaction and knowledge scores. This is somewhat disappointing as this learning strategy is meant to emphasise the application of knowledge, and these higher level application outcomes were not reported in the included studies. We recommend that future studies on TBL look at academic outcomes beyond knowledge retention, such as critical thinking abilities, and non-academic outcomes alike as studies of these outcomes were not found through our searches.

This review included studies investigating undergraduate medical, dental and pharmacy student populations as well as one resident population. Although the heterogeneity of interventions, comparators, designs and outcomes must be taken into account, this review represents the most comprehensive overview of the effects of TBL on health professions education. Most of the included studies took place at the undergraduate level and thus findings of this review are most relevant to these populations. However, the one study of residents by Torralba et al. (2009) is worthy of further examination. This study found a significant improvement in resident in-training exam scores in favour of the TBL groups over the lecture groups for Year 1 residents, but not for Year 2 or 3 residents. We hypothesise that this may be because residents who are further along in their postgraduate training are already exposed to a wide variety of practical applications of their in-class education through their clinical training; whereas Year 1 residents may benefit more from the practical aspects of TBL during their transition to an intensely practical training environment.

In addition to the growing body of evidence suggesting that TBL enables students to achieve knowledge scores as high as, or higher, than traditional teaching strategies, there also exist other potential benefits to TBL not captured in this review. These include a renewed appreciation for group work amongst students, a demand that students take control over their learning and significant reduction in faculty time as compared to many small group learning methods (Searle et al. 2003; Parmelee & Michaelsen 2010).

Conclusions

This review provides the first comprehensive evaluation of the existing evidence on the efficacy of TBL in health professions education. We anticipate it will be of use to educators considering the implementation of TBL in that it demonstrates the potential for TBL to significantly increase knowledge scores. Further, no studies reported a decrease in scores in the TBL-group and we feel this is reassuring to curriculum planners who are looking for active learning strategies that emphasise the application of previously acquired knowledge and to those who have limited numbers of faculty for small group learning activities.

Learner reaction to TBL generally was not higher than the comparator group, even when students' knowledge scores

increased; that is, despite improved performance, students did not prefer TBL to the alternate teaching method. This may be due to the increased student workload and accountability associated with active learning and peer assessment in TBL. More research is required to better understand this discrepancy.

While several areas in need of further research have been outlined, based on the results of this systematic review we would support more widespread implementation of this learning strategy while cautioning curriculum planners to carefully and prospectively consider how they will mitigate potential difficulties in the learners' reactions.

Implications for Research

(1) We recommend that future research on TBL specifically focus on academic outcomes beyond knowledge retention, such as critical thinking abilities and application of knowledge, as studies with these outcomes were not found through our searches and are core to the principles underlying this teaching strategy.
(2) We recommend future research to examine in more detail why learner reaction is mixed and generally not positive compared to evident increases in knowledge retention, as these two outcomes often have significant correlation in other learning strategies.
(3) We find that there is a definite need for more robust primary research to be conducted in TBL with thorough and descriptive reporting. We hope this will allow for statistical meta-analyses or full qualitative meta-synthesis to be performed in this field.
(4) Having established some preliminary evidence for the effectiveness of TBL on improving academic learning outcomes, we recommend that future research work towards drawing specific associations on how and why particular elements of TBL are effective in the way they are.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Notes on Contributors

MIM FATMI is a first-year medical student and a summer studentship recipient. She has completed her Bachelor of Science degree with a major in Biological Sciences and a minor in English. She attended the 2011 Putting Evidence into Practice Systematic Review Workshop by the Alberta Research Centre for Health Evidence at the University of Alberta.

LISA HARTLING, MSc, PhD, is the Director of the Alberta Research Centre for Health Evidence and Director of the University of Alberta Evidence-based Practice Center. In these roles, she oversees approximately 20 staff in the production of systematic reviews, health technology assessments and methodological research for evidence synthesis. She is a reviewer with The Cochrane Collaboration (Acute Respiratory Infections, Anesthesia, Injury, Wounds, Heart, and Infectious Diseases Groups). She has co-authored more than 30 systematic reviews. She was the first author on a systematic review of problem-based learning in medical education and has recently co-authored BEMEs on musculoskeletal Clinical Skills and Audience Response Systems.

TRACEY HILLIER, BSc, MD, CCFP, FRCPC, is a Diagnostic Radiologist in the Canadian Armed Forces and works at the Misericordia Hospital in Edmonton, Alberta. She is also an assistant professor and has recently taken on the position of Assistant Dean of Curriculum for the Undergraduate Medical Education Office at the University of Alberta.

SANDRA CAMPBELL, BA, MLS, AALIA (CP), is a Public Services Librarian and library liaison to the Faculty of Medicine and Dentistry. She is also an expert database searcher and information literacy instructor. She has recently co-authored and provided librarian support and consultation on recent BEMEs on musculoskeletal Clinical Skills and Audience Response Systems.

ANNA E. OSWALD, BMSc, MD, MMEd, FRCPC, is an Assistant Professor, Consultant Rheumatologist and Course Coordinator for the undergraduate musculoskeletal medicine course for preclinical medical students at the University of Alberta and a CanMEDS Clinical Educator for the Royal College of Physicians and Surgeons of Canada. She has a Masters in Medical Education from the University of Dundee. She has co-authored a systematic review on problem-based learning and was the team leader for two other BEMEs on musculoskeletal Clinical Skills and Audience Response Systems.

Glossary

CBGD – Case-based group discussion takes place in small groups and centres around application problems. In health professions education, a team of about 5–15 learners works on a patient case simulating a real-life clinical case.
GRAT – A group readiness assurance test is a multiple-choice quiz taken after an individual readiness assurance test and contains identical questions. It requires discussion and consensus amongst the entire group. Groups are then given immediate feedback after finalizing their answers.
IRAT – An individual readiness assurance test is a multiple-choice quiz administered to ensure that learners have completed and understood the assigned reading. The same test is subsequently administered as a GRAT.
PBL – Problem-based learning is an active learning strategy developed and introduced into health professions education as an alternative to traditional lectures. Learners discuss an evolving case in small groups with the instructor acting as a facilitator. There is an expectation that the group will identify gaps in knowledge related to the case, go away and research around these knowledge gaps and return to apply this information to the case.
SGL – Small group learning is a general form of student-centred learning that takes place in groups of about 5–15 learners.
TBL – Team-based learning is a well-defined active learning strategy developed in the 1970s by Dr. Larry Michaelsen and was implemented in health professions education at the start of the twenty-first century. For the purpose of this review, TBL is defined as a learning strategy that includes the three classic phases: advanced individual preparation, readiness assurance and group application (Parmelee & Michaelsen, 2010). To minimize the heterogeneity of studies reporting use of TBL, we used the above definition in this review. We validated the definition with the literature (Michaelsen et al. 2008), with the founder (Larry Michaelsen) and with a second key expert in this field (Dean Parmelee). The validated definition of TBL includes the three phases above while focusing on learner-

centred education, individual and group accountability, permanence of teams, immediate feedback and a meaningful peer evaluation process (Michaelsen & Parmelee, personal communication, 2011). The nature of the group assignments must be described in detail and must aim to include full cooperation of the team.

Traditional Lecture – Defined as an instructor lecturing to a group of learners (varying in size) in some didactic presentation format. Learner interaction and group discussion is usually minimal in a traditional lecture.

References

Amin Z, Eng KH. 2009. Basics in Medical Education. Hackensack, NJ: World Scientific.

Conway SE, Johnson JL, Ripley TL. 2010. Integration of team-based learning strategies into a cardiovascular module. Am J Pharm Educ 74(2):35.

Davidson L. 2009. Educational innovation in an undergraduate medical course: Implementation of a blended e-learning, team-based learning model. Kingston, ON: Queen's University (Canada). Available from: http://qspace.library.queensu.ca/bitstream/1974/1667/3/Davidson_L_K_200901_MEd.pdf.

Forsetlund L, Bjørndal A, Rashidian A, Jamtvedt G, O'brien Mary A, Wolf F, Davis D, Odgaard-Jensen J, Oxman Andrew D. 2009. Continuing education meetings and workshops: Effects on professional practice and health care outcomes. Chichester, UK: John Wiley & Sons.

Haidet P, O'malley KJ, Richards B. 2002. An initial experience with "team learning" in medical education. Acad Med 77:40–44.

Hartling L, Hamm M, Milne A, Vandermeer B, Santaguida PL, Ansari M, Tsertsvadze A, Hempel S, Shekelle P, Dryden DM. 2012. Validity and inter-rater reliability testing of quality assessment instruments. Rockville, MD: Agency for Healthcare Research and Quality.

Hartling L, Spooner C, Tjosvold L, Oswald A. 2010. Problem-based learning in pre-clinical medical education: 22 years of outcome research. Med Teach 32(1):28–35.

Higgins PT, Green S. (Eds.). 2009. Cochrane handbook for systematic reviews of interventions. Available from http://cochrane-handbook.org.

Kirkpatrick DL, Kirkpatrick JD. 2006. Evaluating training programs: The four levels (3rd ed.). San Francisco, CA: Berrett-Koehler.

Koles P, Nelson S, Stolfi A, Parmelee D, DeStephen D. 2005. Active learning in a Year 2 pathology curriculum. Med Educ 39:1045–1055.

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(11):1739–1745.

Letassy NA, Fugate SE, Medina MS, Stroup JS, Britton ML. 2008. Using team-based learning in an endocrine module taught across two campuses. Am J Pharm Educ 72(5):103.

Levine RE, O'Boyle M, Haidet P, Lynn DJ, Stone MM, Wolf DV, Paniagua FA. 2004. Transforming a clinical clerkship with team learning. Teach Learn Med 16:270–275.

Mennenga H. 2010. Team-based learning: Engagement and accountability with psychometric analysis of a new instrument. Las Vegas: University of Nevada.

Michaelsen L, Parmelee D, McMahon K, Levine R. 2008. Team-based learning for health professions education: A guide to using small groups for improving learning. Sterling, VA: Stylus Publishing.

Morrison A, Moulton K, Clark M, Polisena J, Fiander M, Mierzwinski-Urban M, et al. 2009. English-language restriction when conducting systematic review-based meta-analyses: Systematic review of publish studies. Canadian Agency for Drugs and Technologies in Health.

Nieder GL, Parmelee DX, Stolfi A, Hudes PD. 2005. Team-based learning in a medical gross anatomy and embryology course. Clin Anat 18(1):56–63.

Ogawa RT, Malen B. 1991. Towards rigor in reviews of multivocal literatures: Applying the exploratory case study method. Rev Educ Res 61:265–286.

Parmelee DX, Michaelsen LK. 2010. Twelve tips for doing effective Team-Based Learning (TBL). Med Teach 32(2):118–122.

Pileggi R, O'Neill PN. 2008. Team-based learning using an audience response system: An innovative method of teaching diagnosis to undergraduate dental students. J Dental Educ 72(10):1182–1188.

Prince M. 2004. Does active learning work? A review of the research. J Eng Educ Washington 93:223–232.

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(10 Suppl):S55–S58.

Thomas PA, Bowen CW. 2011. A controlled trial of team-based learning in an ambulatory medicine clerkship for medical students. Teach Learn Med 23(1):31–36.

Torralba KD, Ben-Ari R, Quismorio FP, Boateng BA. 2009. Arthritis and rheumatism conference: American College of Rheumatology/Association of Rheumatology Health Professionals Annual Scientific Meeting, ACR/ARHP 09 Atlanta, GA United States. Conference Start: 2010/11/06 Conference End: 2010/11/11.Conferenc(TRUNCATED), 60, pp. 1356.

Wells G, Shea B, O'Connell J, Robertson J, Peterson J, Welch V, Losos M, Tugwell P. 2000. The NewcastleOttawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analysis [Abstract]. 3rd Symposium on systematic reviews: Beyond the basics, July 3–5; Oxford.

Wiener H, Plass H, Marz R. 2009. Team-based learning in intensive course format for first-year medical students. Croatian Med J 50(1):69–76.

Willett LR, Rosevear GC, Kim S. 2011. A trial of team-based versus small-group learning for second-year medical students: Does the size of the small group make a difference? Teach Learn Med 23(1):28–30.

Zgheib NK, Simaan JA, Sabra R. 2010. Using team-based learning to teach pharmacology to second year medical students improves student performance. Med Teach 32(2):130–135.

Zingone MM, Franks AS, Guirguis AB, George CM, Howard-Thompson A, Heidel RE. 2010. Comparing team-based and mixed active-learning methods in an ambulatory care elective course. Am J Pharm Educ 74(9):160.