ABSTRACT

The Effectiveness of Team-Based Learning: A Meta-Analysis

Sin-Ning Cindy Liu, M.A.

Mentor: Alex Beaujean, Ph.D.

Team Based Learning (TBL) is a pedagogical method developed by Larry K. Michaelsen as a response to the problems of large classes (e.g., low student motivation, low levels of student participation, low class attendance) (Parmelee, 2010). Since its development, TBL has been widely used by educators in many fields and in many countries (Parmelee, 2010; Sweet & Michaelsen, 2007). In this study, a meta-analysis was conducted on existing TBL research. The hypothesis of the study was that the use of TBL will have some effect on student academic outcomes. The moderating effects of country of origin, outcome measure type, education level, and course subject were analyzed. On average, across all studies, TBL seemed to produce better academic outcomes than the comparison pedagogical methods.

The Effectiveness of Team-Based Learning: A Meta-Analysis

by

Sin-Ning Cindy Liu, B.A.

A Thesis

Approved by the Department of Educational Psychology

Terrill F. Saxon, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of

Master of Arts

Approved by the Thesis Committee

A. Alexander Beaujean, Ph.D., Chairperson

Grant B. Morgan, Ph.D.

Wade C. Rowatt, Ph.D.

Accepted by the Graduate School

August 2016

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

DEDICATION

To my mentors, my family, and my family away from home

CHAPTER ONE

Introduction

Team-Based Learning (TBL) is a pedagogical method developed as a response to the problems of large classes (e.g., low student motivation, low levels of student participation, low class attendance) (Parmelee, 2010). In TBL, one instructor oversees a class of students as they work in permanent teams for the duration of a term (Sweet & Michaelsen, 2007). TBL purports to counter the negative effects of having large classes, as students are engaged in active learning even in classes of up to 200 (Haidet et al., 2012). This is accomplished by controlling for social loafing through accountability and a weighted grading system (Michaelsen & Sweet, 2008). Furthermore, the focus of TBL on applying course material to real-life scenarios prepares students for situations they will encounter in their professional experiences (Parmelee, Michaelsen, Cook, & Hudes, 2012).

Since its development, TBL has been widely used by educators in many fields (e.g., medical education, engineering, law, psychology, business, mathematics, statistics, and education) (Haidet et al., 2014; Mennenga, 2012; Stamatel, Bushway, & Roberson, 2013; Sweet & Michaelsen, 2007). TBL has been used by educators in Asia, Europe, the Middle East, and Australia (Burgess et al., 2014; Fatmi, Hartling, Hillier, Campbell, & Oswald, 2013; Haidet et al., 2012; Parmelee, 2010). As of yet, no meta-analysis has been conducted on TBL.

In this study, a meta-analysis was conducted on 38 articles about TBL to examine whether TBL had an effect on student academic outcomes. The mean effect sizes were calculated individually for each outcome measure in each study. Then, moderators were examined to see if they could account for the variability found in the baseline model. The moderators examined in the analysis were country of origin, outcome measure type, educational level, and course subject.

Without accounting for moderators, the average effect of TBL seems to be positive: $g = 0.36$, with a 95% confidence interval of 0.23 to 0.50. Thus, TBL may produce better academic outcomes when compared with other pedagogical methods. Two factors (country of origin and outcome measure) accounted for some of the variability within the data and two factors (educational level and course subject) did not.

CHAPTER TWO

Literature Review

History of Team-Based Learning

Team-Based Learning (TBL) was originally developed by Larry K. Michaelsen in the University of Oklahoma Business School in the early 1990s to alleviate the problems of large classes, low motivation, and low class attendance (Michaelsen, 1992). Michaelsen questioned the effectiveness of lecture-based courses in equipping students with the necessary skills to thrive in a professional environment (Parmelee, 2010). He suspected that there was a disconnection between what students learned in class and what they were expected to do at their jobs in a professional setting. He believed that his students could not fully benefit from learning the course material until they had frequent opportunities to experience and work through the kinds of problems that they would face in the business world, during class (Parmelee et al., 2012).

In 2001, the U. S. Department of Education awarded a Fund for Improvement of Post-Secondary Education (FIPSE) grant to Baylor College of Medicine to incorporate TBL to health professions education. Up until this point, the usage of TBL remained limited in scope. However, this FIPSE grant led to a spike in popularity for TBL use in the medical field (Burgess, McGregor, & Mellis, 2014). TBL workshops were developed and national TBL symposia were formed as a result of the grant and the subsequent increase in TBL usage. The *Team-Based Learning Collaborative* was established as a nonprofit organization in 2003 for the development and practice of TBL (Parmelee, 2010; Parmelee & Hudes, 2012; Thompson et al., 2007).

From 2000 to 2003, the Baylor College of Medicine FIPSE grant led to the adoption of TBL in 10 medical schools in the United States (Searle, Haidet, Kelly, Schneider, Seidel, & Richards, 2003; Thompson et al., 2007). Since then, TBL has further expanded in health education across the United States and in other countries (Parmelee et al., 2012). In 1998, no medical schools were known to use TBL, but by 2014, more than 100 schools worldwide were using the method to some extent (Haidet, Kubitz, & McCormack, 2014). TBL is applicable to health education because physicians are required to work collaboratively with other members in a healthcare team. Exposure to TBL allows students to gain skills that will be useful in making future clinical decisions (Parmelee & Hudes, 2012). Parmelee (2010) expects that “TBL will become increasingly adopted by medical educators, because it addresses so many of the professional competencies and enhances the learning experience” (Parmelee, 2010).

Since its inception, TBL has spread from the United States to many countries, including Japan, South Korea, Singapore, Turkey, Israel, Australia, United Arab Emirates, Oman, India, Austria, and Lebanon (Burgess et al., 2014; Fatmi, Hartling, Hillier, Campbell, & Oswald, 2013; Haidet et al., 2012; Parmelee, 2010).

TBL has been incorporated into many different fields since its beginnings in business education. The increased implementation of TBL in medical education has led to its adoption in medical ethics, neurology, pharmacology, anatomy, evidence-based medicine, ambulatory care, psychiatry, pathology, physiology, nursing education, pharmacy education, psychiatry, and toxicology (Burgess, McGregor, & Mellis, 2014;

Haidet et al., 2014; Mennenga, 2012). TBL has also been extended for use in computer science, economics, electrical engineering, sport psychology, microbiology, criminal justice education, law, marketing, accounting, social work, education, public health, psychology, information management, mathematics, statistics, and communication (Haidet et al., 2014; Mennenga, 2012; Stamatel, Bushway, & Roberson, 2013; Sweet & Michaelsen, 2007). TBL has been utilized at several levels of training: undergraduate, post-graduate and continuing professional education (Parmelee et al., 2012).

Introduction to Team-Based Learning

Team-Based Learning (TBL) is a form of team learning with specific criteria differentiating it from other forms of team, group, cooperative, and collaborative learning. Students in a TBL classroom are organized into permanent small groups for the duration of an academic term. These teams are strategically chosen by the instructor (Sweet & Michaelsen, 2007).

In TBL, the course content is organized into major units of instruction. Each unit of instruction is designed to facilitate the practical application of course material to problems and scenarios similar to those encountered by professionals in the field. Each of these units, or modules, begins with a *readiness assurance process* (RAP) to determine how prepared students are to apply course concepts (Sweet & Michaelsen, 2007). The RAP is followed by *group application* (GA) exercises that challenge students to use course content knowledge to solve problems similar to those that they would encounter in a professional setting (Michaelsen & Sweet, 2008). Figure 1 shows the layout and timeline of a standard team-based learning sequence (Michaelsen & Sweet, 2011).

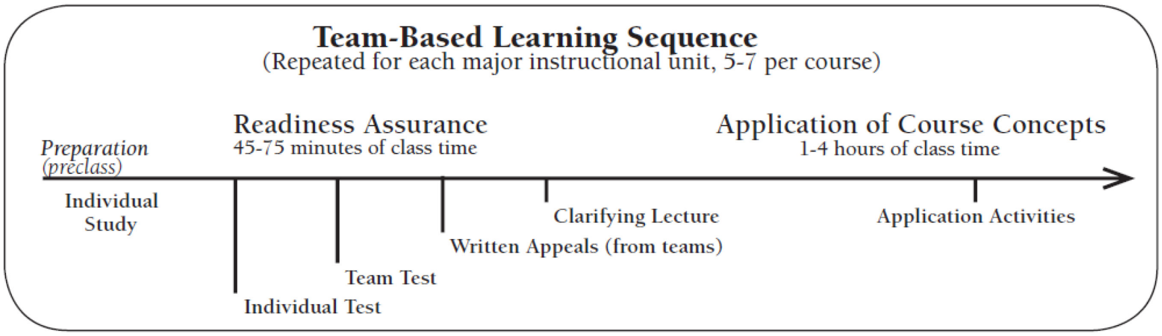

Figure 1. Layout of a standard team-based learning sequence (Michaelsen & Sweet, 2011).

Group Formation

TBL teams need to be formed early in the term using a transparent process to minimize any student perceptions of unfairness (Michaelsen & Sweet, 2008). Classes should be divided into teams of five to seven students, as groups of this size are small enough to work cohesively, but large enough to have the adequate intellectual resources needed to fulfill the demands of TBL (Burgess et al., 2014).

The instructor forms the groups with the goal of evenly distributing member assets (e.g., work experience, previous relevant coursework, different cultural perspectives) and liabilities (e.g. English fluency, negative attitudes towards the course) between the groups (Roebuck, 1998). Thus, the goal is to have heterogeneity within teams (i.e., team members have different skills, assets, and perspectives) and homogeneity across teams (i.e., different teams in a class have comparable group compositions) (Sweet & Michaelsen, 2007). This is essential because heterogeneous teams have more opportunities to discuss and view issues from a variety of perspectives (Michaelsen & Sweet, 2008; Parmelee et al., 2012; Roebuck, 1998). It is important that no pre-formed coalitions (e.g., friends, romantic partners, sorority sisters/fraternity brothers) exist within the teams, to minimize the formation of any cliques or sub-groups within the teams (Michaelsen & Sweet, 2008; Roebuck, 1998).

In TBL courses, students remain in the same teams for the duration of the course. Because the groups are heterogeneous, it takes longer for members to solidify as a team than those in homogeneous groups. Keeping teams together for the duration of an entire term maximizes the potential effectiveness of the teams (Michaelsen & Sweet, 2008; Parmelee & Hudes, 2012; Sweet & Michaelsen, 2007).

Readiness Assurance Process

Every TBL module begins with a readiness assurance process (RAP), which has five key outcomes: (a) effective and efficient content overage; (b) development of teams and teamwork skills; (c) learning the value of considering input from diverse sources; (d) development of self-study and life-long learning skills; and (e) optimal use of class time (Farland et al., 2013; Parmelee & Michaelsen, 2010). The RAP occurs in four parts: (a) an individual readiness assurance test (iRAT), (b) a team or group readiness assurance test (tRAT/gRAT), (c) written appeals, and (d) instructor feedback (Michaelsen & Sweet, 2011).

iRAT and tRAT

The iRAT and tRAT/gRAT are not specific instruments that are administered to the students. Rather, they are instructor-created tests to examine how well students understand the information within a specific module (Michaelsen & Sweet, 2011). As students are expected to study the material before class, the iRATs serve as a measure of student preparation for class (Roebuck, 1998).

For the iRAT, instructors typically use short multiple-choice tests that cover the material in the assigned readings. These tests are usually 10-20 questions long (Farland et al., 2013; Parmelee & Hudes, 2012). Since part of the TBL process requires students to solve application problems as a team, instructors design iRAT questions to focus on the concepts that students need to master in order to solve the application problems (Parmelee et al., 2012). As a tool for instructors to gauge the students' relative levels of understanding, the iRAT allows the instructor to determine whether the students can be reasonably expected to use course concepts in application exercises (Roebuck, 1998). Moreover, iRATs allow students to be responsible for their own learning through independent study prior to class meetings (Michaelsen, 1992).

After individuals submit their iRAT responses, teams convene to retake the test, except this time the students complete the test as a group. This re-testing is referred to as the tRAT (Sweet & Michaelsen, 2007). It is important that the iRAT precedes the tRAT to ensure that students do not solely rely on the team to mask a lack of individual preparation (Stamatel et al., 2013). When students take the tRAT, the teams engage in peer instruction as team members take turns facilitating the discussion by asking elaborative questions, much as an instructor would. This co-regulation of learning by team members eliminates the need for continual instructor supervision throughout the learning process (Michaelsen & Sweet, 2011). Researchers have reported that the language used by students during peer teaching is often more accessible than the language used by instructors (i.e., the students use words that their team members find easier to understand, when compared with the language used by instructors in lectures). This may be because students tend to explain course concepts using simpler vocabulary, as opposed to the more technical language that instructors often use. By using more colloquial explanations, students may successfully engage team members in effective peer instruction (Haidet et al., 2012).

During tRAT discussions, students share what they learned through independent study as each team member explains their rationale for supporting one answer over another. As conflict is uncomfortable, brand-new teams will generally make their final answer choices through voting. With time, students tend to learn that voting is not the most effective way to reach an answer on the tRAT. Consequently, they often transition to choosing an answer through team discussion and coming to a consensus (Parmelee et al., 2012). Learning is promoted in the tRAT discussions as students need to articulate their reasons for their chosen answer (Michaelsen, 1992). The input that students receive from their teammates serves to broaden their content knowledge. As students explain the logic for their answers to the tRAT questions using information from the material, the discussions clear up misconceptions about course concepts (Roebuck, 1998).

Groups often learn that it is more harmful than beneficial to have only a few group members dominate team discussions (Michaelsen 1992; Roebuck, 1998). Students dominating team discussions will inevitably get an answer wrong at some point. When this occurs, these students become more cautious and encourage the quieter, more passive team members to speak up, especially if these team members answered questions correctly before (Parmelee et al., 2012). Thus, more vocal members tend to talk less and listen more than they would in other group settings, and quieter members are encouraged to participate in the discussion to a greater degree than they otherwise would (Michaelsen, 1992; Roebuck, 1998).

In an analysis of nearly 1600 teams over a period of 20 years, Michaelsen, Watson & Black (1989) found that over 99.9% of the teams outperformed their own best member by an average of almost 11%. In the majority of the classes they studied, the lowest team score in the class was still higher than the best individual score of the entire class (Michaelsen & Sweet, 2008). These results are powerful incentives for fostering group cohesiveness. Team members tend to become more supportive of each other through the course of a semester as the increased familiarity with team members allow students to become better at explaining concepts to others on their team (Michaelsen & Sweet, 2008; Roebuck, 1998). This cohesiveness often leads to the development of group norms, as unspoken expectations for student behaviors form. Teams expect their members to maintain good class attendance and come to class prepared to participate in group discussions. As these norms form, instructors find that they often do not need to enforce attendance or pre-class preparation (Roebuck, 1998).

Immediate Feedback-Answer Technique Forms

Michaelsen, Knight, and Fink (2004) recommended that instructors conduct the tRAT using self-scoring answer sheets known as the *immediate feedback-answer technique* (IF-AT) form. An IF-AT form looks similar to a lottery scratch-off card, with boxes covered by opaque film. To use an IF-AT form, students scratch off the film covering one of the boxes to indicate their answer choice for a question. If their choice is correct, a mark (e.g., a star in Figure 2) is uncovered once the film is removed. Full credit is rewarded to the team if the mark is revealed on the team's first answer attempt. The team can continue scratching an IF-AT form until they find the mark. A score penalty is issued, however, with every failed scratch attempt. For example, a team may receive one point if they answer the tRAT question correctly on the first try, half a point if they answer correctly on the second try, or a quarter-point if they answer correctly on the third try. Thus, an IF-AT form awards credit for partial knowledge by allowing teams to make multiple attempts at finding the answer (Michaelsen & Sweet, 2008).

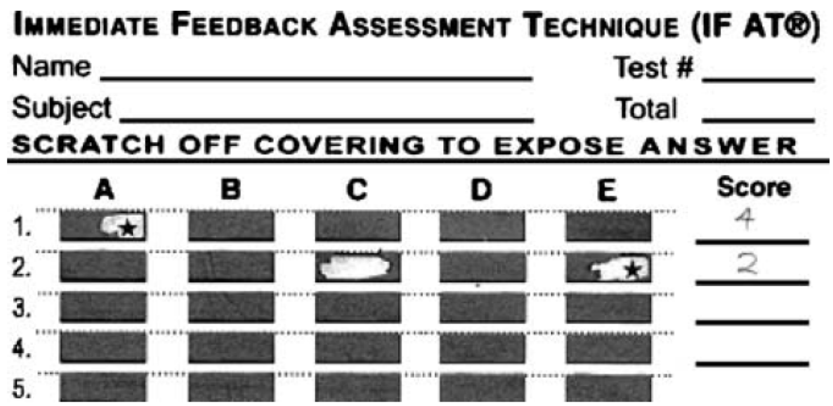

Figure 2. The Immediate Feedback Assessment Technique (IF-AT Form) with correct answers uncovered (Michaelsen & Sweet, 2007).

Practitioners of TBL have discovered that the utilization of IF-AT forms for the tRAT often discourages outspoken students from dominating group discussion (Michaelsen & Sweet, 2008; Sweet & Michaelsen, 2007). Since the grades of all team members are affected by the tRAT scores, team members tend to be motivated to join the discussion in hopes of finding the correct answers to the test. The real-time feedback provided by the IF-AT form motivates the teams to work effectively together without instructor input. Michaelsen and Sweet (2008) asserted that immediate feedback is the "most effective tool available for promoting both concept understanding and cohesiveness in learning teams" (p. 18).

While IF-AT forms are not a requirement for TBL, it is vitally important for teams to receive immediate feedback on each of their answers during the tRAT process, as feedback is a powerful catalyst for learning (Hattie & Timperley, 2007; Michaelsen & Sweet, 2011). However, besides the IF-AT form, there are other methods to facilitate tRATs. An alternative to the IF-AT form that provides immediate feedback to students is the audience response system "clickers" (Parmelee & Hudes, 2012).

Another alternative is for instructors to assign each tRAT question a number of points (e.g., a question with four answers is worth four points) and allow teams to distribute the points to one or more answer choices depending on how confident they are of the answer. For example, on a question with four answer choices, a team could assign 1 point to A and 3 points to B. If the answer is A, then the team receives one point. If the answer is B, then the team receives three points. If the answer is C or D, then the team receives no points. While this third method awards some credit for partial knowledge by allowing students to distribute points to multiple answer choices when they are not fully confident in any one choice, it does not provide the immediate feedback of the IF-AT or "clickers" (Farland et al., 2013).

As teams encounter problems during discussions (i.e., when they get an answer wrong on the tRAT), members can assess the effectiveness of the team's problem solving and communication strategies and amend their approach as necessary (Haidet et al., 2012). The immediate feedback from the tRATs can help improve the decision-making processes of teams as they work to answer questions through consensus-building discussions, making teams more effective and efficient through the course of a term (Parmelee et al., 2012). The promise of a reward for partial knowledge tends to encourage teams to continue discussing a tRAT question even if their initial response is incorrect. When a team encounters a wrong answer, they explore why they missed the question and continue to debate the merits of the remaining choices, motivated by the partial credit they can receive (Brandler, Laser, Williamson, Louie, & Esposito, 2014).

Appeals

At the conclusion of the RATs, teams (but not individuals) can appeal any questions that they missed by making a convincing case that: (a) their answer can be considered correct, (b) the item requires revision, or (c) the assigned reading material inadequately prepared them to answer the item (Sweet & Michaelsen, 2007). To appeal an item, teams need to provide a clear and usable rewrite of a poorly worded or ambiguous question or submit a written explanation for why their answer choice was as appropriate as the "best" choice indicated by the instructor (Parmelee et al., 2012).

To submit an appeal, teams must support their arguments with evidence from the course materials. This process must happen immediately following the tRAT. If the instructor grants the appeal, only teams that successfully appealed the test item are awarded points (Sweet & Michaelsen, 2007). These points are added to both the iRAT and tRAT scores of the team members (Farland et al., 2013). The appeals process can act as a catalyst for team cohesion as each team autonomously works on its own behalf, since no team receives the benefits of another team's work. As the teams are challenged to write well-constructed appeals, they become more cohesive as they work together to organize their ideas into persuasive arguments (Roebuck, 1998).

The appeals process is an important part of RAP because it is an opportunity for students to review the material with which they were least familiar. Instructors report that students tend to learn more from appealing items they got wrong than from confirming items they got right, likely because teams need to have sufficient evidence to support the appeals (Michaelsen & Sweet, 2008). The appeals process is an opportunity for students to channel their disappointment over missing an item into a review of difficult material and troublesome concepts (Michaelsen, 1992; Roebuck, 1998). By requiring teams to submit appeals in writing, students practice systematically creating their responses using coherent, persuasive, and sound logical arguments (Michaelsen, 1992; Roebuck, 1998). In addition, the appeals process affords instructors the opportunity and time to evaluate the appeals privately after class and respond thoughtfully (Michaelsen, 1992).

Lecture

The last phase of the RAP is a lecture by instructors specifically focusing on information that they believe students do not fully understand. After the tRAT and appeals processes, instructors are aware of which topics challenge and confuse the students most. The lecture is an opportunity for instructors to discuss the most challenging aspects of the assigned materials, tailoring the scope of the lecture to the specific material that the students had difficulty understanding (Michaelsen & Sweet, 2008). The RAP allows instructors to assume that individual study and team discussions adequately covered the material in the RAT items that students answered correctly. (Farland et al., 2013; Michaelsen & Sweet, 2011; Stamatel et al., 2013). Through bypassing material that students have already mastered, instructors can spend more class time on application-oriented activities that occur after RAP in the TBL module sequence (Sweet & Michaelsen, 2007).

According to Michaelsen and Sweet (2011), the feedback from the iRAT and tRAT primes students to be more alert during TBL lectures than they would be in standard lecture classes. Students know which parts of the material remain confusing to them and listen actively to hone in on the information in the lecture that could clear up these misunderstandings.

Group Application Exercises

After the RAP, the majority of each module is spent on assignments and exercises that allow students to deepen their understanding of the material through practical applications of their knowledge (Michaelsen & Sweet, 2008). These activities are referred to as *group application* (GA) or *team application* (tAPP) exercises (Parmelee, 2010; Parmelee et al., 2012). While RAT questions assess lower levels of Bloom's taxonomy (i.e., remember, understand, and apply), instructors design GA exercises at higher levels of the taxonomy (i.e., analyze, evaluate, create) (Bloom, 1974; Farland et al., 2013).

Some examples of GA exercises are case studies—scenarios and vignettes of problems similar to those that students will face in a professional setting. In GA exercises, students can only arrive at solutions through careful deliberation and debate over information from the course content, thus applying their knowledge to a practical situation (Parmelee et al., 2012; Roebuck, 1998; Stamatel et al., 2013). Each group completes the GA exercises and presents their results to the entire class at the conclusion of each exercise. Thus, the GA component of TBL includes both inter-group and intra-group discussion (Michaelsen & Sweet, 2008).

Michaelsen and Sweet (2008) established some criteria for instructors who create GA exercises. GA tasks that promote learning and team development require group interaction. It should be impossible to complete GA assignments by dividing up the workload and having individuals execute project components independently. GA questions need to engage students' higher-order levels of thinking and involve complex concepts from course materials in order to facilitate effective interaction between team members. Decisions on answers to GA problems should be challenging and require intra-team discussions. As GA questions require an answer within a short time period, team members do not have time to research the topic independently or split the work after class. Lastly, GA assignments should be presented using a basic format (e.g., choosing a letter that represents a decision, presenting information through charts or graphic organizers) so that students are able to mainly focus on the content and logical soundness of their answer without worrying as much about the presentation of the material (Michaelsen & Sweet, 2008).

Practitioners of TBL believe that in order to utilize course concepts in real-life situations, students need to have active practice in applying the concepts through GA exercises (Haidet et al., 2014). By examining complex problems or case studies, students can see how course content is related to real-life scenarios. This process should minimize the idea that the information students are learning is too abstract or removed from actual practice (Michaelsen & Sweet, 2011).

Besides utilizing higher order thinking skills, GA exercises also foster the development of non-cognitive skills such as teamwork and communication (Farland et al., 2013). GA exercises are opportunities for teams to solve ambiguous, complex problems simulating real-world conditions (Haidet et al., 2012). When students can see how they can apply course concepts to realistic problems and scenarios, they tend to be more motivated to learn and participate in the course (Roebuck, 1998). The team structure is vital to the GA portion of TBL, since teams are more equipped to meet the challenge of addressing problems that are too difficult and complex for individuals to complete independently (Michaelsen, 1992).

Designing a Team-Based Course: Backwards Design

Designing a TBL course requires instructors to "think backwards" by planning the course around what they want students to be able to do at the conclusion of the course. This process is called *backwards design* (Michaelsen & Sweet, 2008). RAT questions need to be appropriately tailored so that students feel adequately prepared for the GA exercises, while GA tasks should require mastery of certain skills and concepts from the material (Parmelee, 2010; Stamatel et al., 2013).

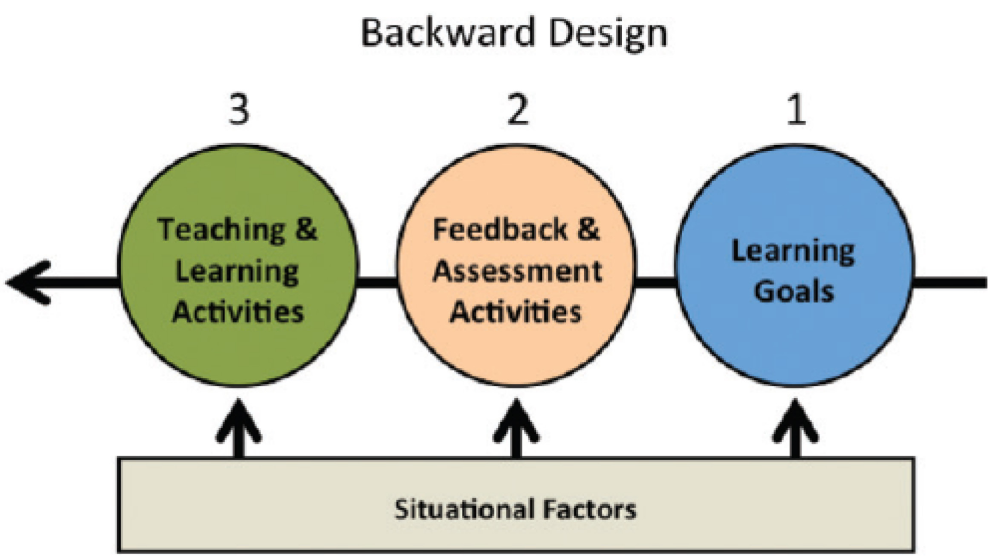

Figure 3. Backwards design (Parmelee et al., 2012).

Learning Goals

The first step of backwards design is for the instructor to identify one to three instructional goals for each class (Stamatel et al., 2013). These learning goals need to be clear, specific and meaningful, and answer the question, "What do I want my students to be able to do at the end of the session that they could not do before?" Instructors are encouraged to use action verbs to specify how this mastery should be achieved (Parmelee et al., 2012).

Designing GA Exercises

Once instructors identify the learning goal of the course, they design GA exercises to help students achieve these outcomes through practice, feedback, and assessment. The best GA cases are often complex, so that students cannot simply find the answer by looking up the answer in course materials, but must arrive at the conclusions through deliberation, reason, and logic. At the same time, these cases should focus on major overarching concepts instead of details, to lessen the risk of leaving gaps in student learning. GA problems also need to be authentic problems similar to those that students will encounter in professional settings (Farland et al., 2013; Parmelee et al., 2012).

Designing RATs

After the instructors design the GA exercises, they can create RATs that prepare students for the GA tasks. The RATs should consist of multiple-choice questions that are well-constructed, of comparable quality to course or licensing exams, and focus on overarching course concepts (Parmelee et al., 2012).

Questions to Guide Backwards Design

Michaelsen & Sweet (2008), Parmelee (2010), and Roebuck (1998) suggested that instructors ask the following questions during backwards design to develop useful RAT questions and GA exercises:

- What will students need to know in order to fulfill learning objectives?
- What knowledge will students need to make decisions when solving problems?
- What criteria separate a well-made decision from a poorly-made decision using this knowledge?
- What do we want our students to be able to do with this information two or three years from now?
- What should students be able to accomplish when they have completed this unit of instruction?
- How can the instructor assess what knowledge students already have in order to build on existing knowledge rather than waste time dwelling on what they already know?
- How can one determine whether or not students can effectively use their knowledge?

The 4 Ss

According to Michaelsen, in order to maximize the effectiveness of TBL, all assignments should be characterized by the 4 Ss: *Significant problem, Same problem*, *Specific choice*, and *Simultaneous report*. The 4 Ss are used at all three stages of learning in TBL: pre-class individual preparation, intra-group discussions, and full class inter-group discussions (Michaelsen & Sweet, 2008).

Significant Problem

Students should work on problems, cases, and questions that demonstrate the usefulness, applicability, and relevance of the concepts in the module. These concepts should be interesting to the students so that they are engaged in answering the problems. Backwards design is especially helpful for identifying GA questions that are meaningful to students (Michaelsen & Sweet, 2008; Michaelsen & Sweet, 2011).

The GA problems that instructors give students should represent the kinds of issues they could face in the workplace or in future coursework. As such, the answers to these problems should not be found in any one source. Instead, students can only find the answers through discussion (Parmelee et al., 2012).

Same Problem

Instructors should give all students the same materials to study during the pre-class preparation period. Furthermore, all teams should work on the same problems in GA. Working on the same tasks while using the same material makes students more likely to care about the conclusions and rationale of other teams—and more carefully defend the logic of their own decisions—than if they worked on different problems (Michaelsen & Sweet, 2011).

Although some instructors may be tempted to assign different questions to different groups in order to cover a wider range of information, this tactic often backfires because students will not be motivated or engaged in the work of other groups (Parmelee et al., 2012). In such a scenario, students are less likely to pay attention when other groups present their findings, eliminating inter-team accountability and minimizing inter-team discussion. Furthermore, students may not feel prepared to present a credible challenge to another team's conclusions if they did not work on the same question. In TBL, considerable inter-team debate and discussion is generated if teams select different answer options for GA questions (Michaelsen & Sweet, 2011; Parmelee et al., 2012).

Specific Choice

All GA problems should be presented with clear alternative decisions. Students should be able to explain why they chose their answer and defend their conclusion with evidence from class materials. Thus, students not only make a decision about a topic, but must also explain how they arrived at that choice (Michaelsen & Sweet, 2011). These GA questions should not have obvious answers but should force teams to discern between several equally plausible answers, select the most appropriate one, and defend their decision with information from the materials (Michaelsen & Sweet, 2008; Parmelee et al., 2012).

Each team makes a choice to the GA questions after thorough intra-team discussions using a simple format for displaying the teams' answer choices, so that the answers of different groups can be easily compared (Parmelee et al., 2012). At most, teams should prepare a one page justification of their answer (Burgess et al., 2014). It is more important that teams are prepared to explain how they excluded the alternate solutions to the question (Parmelee & Hudes, 2012).

Simultaneous Report

Sometimes reporting effects (e.g., primacy effects, answer drift) occur when student responses are given in a sequential (i.e. non-simultaneous) order. The phenomenon of answer drift occurs when an initial response to a question influences subsequent responses to the same question. Those who report their answers after the initial report have a tendency to change their answers to fit the perceived majority view: they want to fit in with the class consensus. This problem is especially marked if the first response was incorrect and students with answers that are more logically sound than the first response feel pressured to change their perspective and also report incorrect findings (Michaelsen & Sweet, 2008).

Answer drift is a common side effect of sequential reporting and limits the display of different perspectives on a certain topic. Later reporters want to downplay their differences from previous reporters. A class consensus—which may not even be correct—often emerges. This consensus is overly influenced by the stance taken by those who provided the first few answers. This phenomenon diminishes potential discussion or debate that could have stemmed from having a large array of answers. The full spectrum of potential answers is not reported due to the students' desire to belong to the majority opinion (Michaelsen & Sweet, 2008).

Simultaneous report is a means to balance this problem. By requiring all groups to simultaneously report their answers to a GA question, answer drift is eliminated. Teams are forced to choose a stance and defend their position, regardless of how other teams respond. This method of reporting provides a greater potential for disagreement between groups, which leads to in-depth, content-rich discussions and debates. As teams collaborate to defend their selected position, intra-team discussions facilitate the development of group cohesion (Michaelsen & Sweet, 2008). The debates that occur at this stage are generally lively and well-constructed, as the students are very familiar with the material (Stamatel et al., 2013). Three methods of simultaneous reporting are as follows: (1) graphic organizers; (2) concept maps; and (3) holding up color-coded or lettered cards in response to answer choices (Michaelsen & Sweet, 2011).

By simultaneously reporting answers, all teams receive immediate feedback on where they stand in relation to other teams (Parmelee et al., 2012). This provides a foundation for productive discussion as students see how their teams' answers compared and contrasted with those of other teams (Farland et al., 2013). Students engaged in intra-team collaboration as team members work together to defend their team's point of view (Haidet et al., 2012). Teams are accountable for reaching a consensus and being prepared to defend their decision (Parmelee et al., 2012).

Potential Benefits and Problems of Team-Based Learning

Large Classrooms

TBL is a pedagogical method shown to compensate for some of the common disadvantages of large classes. Instructors are doubly challenged as school administrators advocate for increasingly larger classes in order to receive more revenue while simultaneously demanding that instructors teach students in ways that are active, engaging and promote positive learning outcomes (Parmelee et al., 2012). In large classrooms, students are able to remain anonymous, which often results in an apathetic attitude towards the course. Instructors are frustrated as student motivation and preparation decrease, but external pressures for accountability from the administration build (Stamatel et al., 2013).

Michaelsen (1992) argued that TBL may be the only means of channeling the higher-level cognitive skills of students in large class settings (Michaelsen, 1992). Using TBL, it is possible for a single instructor to manage and provide real-time supervision and content expertise to a class of up to 200 students operating in small-group learning and problem-solving tasks (Haidet et al., 2012). Groups develop into effective, self-managed learning teams, eliminate anonymity and build accountability (Michaelsen &

Sweet, 2008; Stamatel et al., 2013). TBL is also a cost effective method for large-scale student engagement, as only one instructor is needed per course (Parmelee, 2010).

Attendance

Faculties in many universities have become frustrated with student attendance levels in recent years. As more lectures are recorded and broadcasted online, fewer students attend lectures (Parmelee et al., 2012). However, instructors in TBL courses rarely worry about attendance problems (Michaelsen, 1992). Studies of TBL report increased levels of attendance and engagement (Shankar & Roopa, 2009). The use of a team folder to track the numbers of total absences and unexpected absences for every class period facilitates the formation of norms of class attendance and preparation. Students do not want to let their team members down, as social loafing can be penalized in TBL (Michaelsen & Sweet, 2008). Furthermore, the cost of missing classes in TBL is higher than the cost of missing classes in a lecture based course. A student cannot simply ask a classmate for notes on the RAP or GA exercises, since much of the learning in TBL occurs during small group interactions and discussions. Students who miss TBL sessions often find themselves struggling to catch up and will rarely continue to miss class (Stamatel et al., 2013).

Teamwork

Synergy was calculated by Watson, Michaelsen, & Sharp (1991) by dividing the amount by which the team outscored its best member (team score - best member score) by the amount it was possible for the team to outscore its best member (total possible points - best member score). They found a general trend of individual scores increasing over the course of a term as students became more familiar with the demands of a TBL course, which in turn increases the difficulty for a team to outscore its best member. However, synergy ratios increased over the course of a semester, indicating that team scores improved even more than individual scores did (Sweet & Michaelsen, 2007). This finding supports the idea that instructors can pose questions and exercises that would overwhelm individual students during GA, as the collective team should have adequate resources to complete these tasks (Michaelsen & Sweet, 2011; Roebuck, 1998).

Over the course of a term, instructors often see a shift in the teamwork style of a team from compromising to problem solving (Sweet & Michaelsen, 2007). As these interactions patterns change, team cohesion increases as students develop the skills they need to work as productive team members (Michaelsen, 1992; Michaelsen & Sweet, 2008). Through TBL, students develop interpersonal skills that allow them to work productively in their teams: to listen, communicate, solve problems, make decisions, resolve conflicts, and manage time and resources (Roebuck, 1998).

Experience working in teams teaches group members that the group is collectively more intelligent than any given individual. As the term progresses, students find that their confidence in the abilities of the team has increased, which leads to an increase in the students' confidence in their own abilities to master class content and participate meaningfully in team exercises and discussions (Stamatel et al., 2013). As students work together in a team, increased familiarity with team members allows students to become better at explaining concepts so that their team members can understand them (Michaelsen & Sweet, 2008). Groups also motivate individuals to attend class and prepare for group work as team norms of attendance and preparation develop.

Students do not want to be left behind by their teams, which motivates them to work harder (Michaelsen, 1992; Roebuck, 1998; Tucker & Brewster, 2015).

Teams provide a constant source of feedback for students and allow them to monitor their progress throughout the course of the term (Michaelsen, 1992; Roebuck, 1998). TBL is a unique pedagogical method because students are always aware of whether or not they understand course content, due to the amount of feedback they receive (Parmelee et al., 2012). Teams also provide a social network that benefits students who are in need of social support (e.g., international students, nontraditional students, or students at risk of dropping out) and might not receive the same level of support in a lecture setting (Michaelsen, 1992).

Feedback

Immediate feedback encourages intra-team and inter-team competition, which promotes the acquisition and retention of knowledge. Students are aware of their relative standing in comparison with their peers in terms of understanding course content and applying content knowledge, because of feedback on individual and team performances (Burgess et al., 2014). Immediate feedback also impacts group development (Michaelsen & Sweet, 2008). In TBL, feedback occurs at multiple levels. The iRAT provides feedback on the student's individual pre-class preparation. The tRAT and GA exercises provide peer feedback as the students discuss and debate within their group. Answer reporting for GA exercises provides feedback from other teams and the instructor (Stamatel et al., 2013).

During team discussions and GA exercises, students receive immediate and unambiguous feedback from their teammates. The impact of such feedback is immediate.

Michaelsen & Sweet (2008) argue that peer feedback in TBL may be more beneficial to a student than a one-on-one relationship with the instructor since students are exposed to high volumes of immediate peer feedback during team interactions. Immediate feedback allows students to engage with their peers and with course content (Farland et al., 2013). Students in TBL benefit from having personalized, individual corrective instruction from their team on an ongoing basis for the duration of a term, something students in other pedagogical methods do not have access to (Michaelsen, 1992).

Peer evaluation is another source of feedback in TBL courses and allows students to provide meaningful and honest feedback to peers within a safe environment (Haidet et al., 2012). Peer evaluation systems need to be capable of: (1) accommodating teams of different sizes; (2) accurately reflecting the work of team members; (3) making a significant impact on the course grade (Farland et al., 2013). There are two kinds of peer evaluations: formative and summative. Formative assessments occur in the middle of a term and help students improve their performance as team members and develop interpersonal and team skills. Summative assessments occur at the end of a term and are used by the instructor in the grading process (Cestone et al., 2008).

During peer evaluation, each team member makes an assessment of every other member's contribution to the team's success. This information is then used as part of the grade weighting process (Cestone et al., 2008). While peer evaluations are important, they should not be conducted too often, since frequent evaluations may disrupt team development by establishing a dominant team member early in the semester. Students who participate in frequent evaluations may also experience "survey fatigue". Usually one formative and one summative assessment per term is sufficient (Cestone et al., 2008;

Farland et al., 2013). Instructors should provide guidelines for students to help them provide helpful feedback to their peers (Parmelee et al., 2012). These guidelines can help students concretely understand what constitutes helpful feedback and how to structure effective feedback (Cestone et al., 2008).

Four methods of peer evaluation have been suggested in the TBL literature (Cestone et al., 2008; Farland et al., 2013):

- Michaelsen method: students assign a score to each teammate based on the extent to which that teammate contributed to the team's overall performance, while being forced to make differentiations among peer performances. For example, in a six-person team, 50 points are given to each student to divide among the five team members (excluding self) with a minimum possible score of 7 and maximum possible score of 13. Students are forbidden from assigning the same score to all team members. The overall score for each individual is calculated by summing the scores he or she received from the other team members.

- Fink method: students assign up to 100 points for each team member, depending on the team member's contributions to the team. Each member receives a peer score from the sum of the points they are awarded by the other team members. This peer score is multiplied by the student's mean tRAT score to arrive at an adjusted group score. There is no required differentiation of points, so students may assign full points to each group member.

- Koles method: comprised of anonymous quantitative and qualitative feedback sections. The quantitative section has ratings on cooperative learning, self-directed learning, and interpersonal skills. The qualitative questions ask students to identify the most valuable contributions an individual makes to the team and the most important thing that individual could do to help the team more effectively.

- Lane method: students are involved in the creation of instruments and procedures that are used to collect quantitative and qualitative peer evaluation data. This data is used to provide feedback and grade input for team members.

Instructors also provide important feedback in the TBL process. Students are more likely to ask an instructor a question as he or she is passing by in the TBL classroom than in front of the entire class in a lecture-based course (Stamatel et al., 2013). Instructor feedback occurs when instructors facilitate GA discussions, during tRAT appeals, and in RAP lectures (Michaelsen & Sweet, 2008; Sisk, 2011). Instructors in a TBL course are usually aware of whether their students are struggling with the material, based on the results of the RAP and GA exercises, and tailor their instruction accordingly (Parmelee et al., 2012; Stamatel et al., 2013).

Facilitation

TBL is unique when compared with other active learning pedagogical methods because a single instructor can facilitate a course, regardless of the number of students in the class. The instructor is the content expert and provides a framework and guidelines for TBL sessions. The facilitator prompts teams to verbally justify their answers during inter-group discussions, guide the debate to achieve the learning objectives identified in backwards design, and use simultaneous reports in GA as a catalyst for facilitating inter-team debates (Farland et al., 2013; Parmelee et al., 2012). Ultimately, the job of a TBL facilitator is to design and manage the course in such a way that students can master the material. The instructor defines and identifies course goals, establishes acceptable performance standards, and develops group assignments and team activities that are designed to help students master essential concepts (Roebuck, 1998).

Relevance

One major difference between TBL and lecture-based instruction is that students can experience the relevance of course material during a TBL course. Students do not need to wait for opportunities to apply their knowledge in the future, since the in-class GA exercises mirror scenarios they could face in their professional fields (Michaelsen & Sweet, 2008). The focus on application ensures that the students interact actively and analytically with the course information, based on reflections of how the material would be useful in the future (Stamatel et al., 2013). TBL has an emphasis on solving problems using data and information that is as authentic as possible, so students can experience what it is like to partake in the decision-making processes they will regularly encounter in their future professional experiences (Parmelee et al., 2012).

Effectiveness

Practitioners argue that when TBL is conducted correctly, with all of the proper preparations and procedures, the academic outcomes of TBL are equivalent or improved in comparison to lecture-based or more traditional small-group formats (Parmelee et al., 2012). In a systematic research review conducted by Fatmi et al. (2013), no studies reported a decrease in scores in TBL courses, when compared with courses using other pedagogical methods (Fatmi et al., 2013). In terms of effectiveness, the TBL classroom provides a safe environment for students to hone the practical skills they need for employment as they develop professional competencies such as interpersonal communication, teamwork, and constructive feedback skills (Parmelee & Hudes, 2012; Roebuck, 1998).

TBL has been shown to display increased test performance (Koles, Stolfi, Borges, Nelson, & Parmelee, 2010), increased retention of course material (Michaelsen & Sweet, 2011), decreased student failure rates (Parmelee & Hudes, 2012), and increased National Board of Medical Examiners shelf examination scores (Levine, O'Boyle, Haidet, Lynn, Stone, Wolf, & Paniagua, 2004; Thompson et al., 2007). Studies show that weaker students benefit to a greater extent from TBL than from other pedagogy models, without slowing down the progress of their more capable peers (Parmelee & Hudes, 2012; Sisk, 2011). Students are trained and motivated to become life-long learners equipped to adequately assess research and data in their fields, and develop confidence in their abilities to both understand and apply relevant information to real-world settings (Michaelsen, 1992; Parmelee et al., 2012).

Satisfaction with Team-Based Learning

The implementation of TBL changes the teaching experience for instructors and redefines their roles as educators. The instructor is no longer the creator of knowledge, but is instead the facilitator of knowledge creation in students. Instructors often noted that by the end of the term, students began to feel more like colleagues and co-creators of knowledge, instead of "empty vessels" into which instructors pour information (Michaelsen, 1992; Stamatel et al., 2013).

Adopters of TBL have reported a reduction in student apathy, increased attendance, more critical thinking, rich in-class discussion, and improved performance in their classes (Stamatel et al., 2013; Thompson et al., 2007). Students tend to be more enthusiastic about the material in TBL courses than in lecture-based courses (Michaelsen & Sweet, 2008). Instructors like the increased level of student engagement in TBL settings (Tucker & Brewster, 2015). Furthermore, instructors have shown interest in the development of professional skills—leadership, communication, and teamwork—in students using TBL (Burgess et al., 2014).

Studies of student satisfaction with TBL have shown mixed results. While some studies report high satisfaction with TBL, other studies find no statistically significant differences in satisfaction ratings of TBL when compared with other teaching methods

(Beatty, Kelley, Metzger, Bellebaum, & McAuley, 2009; Sisk, 2011; Thompson et al., 2007; Tucker & Brewster, 2015).

It has been found that students with early exposure to TBL and repeated experiences with TBL displayed an improvement in attitudes towards TBL (Thompson et al., 2007). Students who prefer TBL like the active and collaborative approach (Burgess et al., 2014). Students reported satisfaction from the development of a social support system in their team and from completing difficult but meaningful assignments (Roebuck, 1998).

Student Accountability

Accountability is an important factor built into the TBL process at both the individual and the team levels. Students are accountable to their teammates and the instructor for their individual preparation work (e.g., studying the reading materials before class, being prepared for the iRAT and tRAT, showing up to class). Teams are held accountable to the instructor for the work produced by the team as a whole (e.g., the tRAT, well-written appeals, logically sound arguments for answers to GA exercises) (Michaelsen & Sweet, 2008). Overall, the RAP is a good motivator for student accountability. By testing the knowledge of individuals and of the team, students come to class prepared for the sake of both their individual and team grades (Burgess et al., 2014).

Individual students are accountable to their teammates because lack of preparation is difficult to conceal during the tRAT and GA discussions (Michaelsen & Sweet, 2008). Student accountability in TBL accounts for increases in attendance, pre-class preparation, and in-class participation, since students are committed to their teams (Roebuck, 1998). Students in TBL are aware that their peers know how prepared have been for class and their grades are directly affected by evaluations from these same peers (Michaelsen & Sweet, 2008).

Michaelsen (1992) proposed a way to tangibly promote student accountability by providing teams with constant data about how their members are doing in the course. Each team has a folder with a form that is updated at every class session to track each team member's performance. The tRAT scores for the team, iRAT scores for each individual student, total number of member absences, and number of unexpected student absences are recorded. The names of absent members are not recorded and the iRAT scores are labeled using the students' school identification numbers in lieu of their names. However, the records are public enough to support the development of group norms that encourage preparation, attendance, and participation (Michaelsen, 1992).

Teams are motivated to produce high quality work because all teams are working on the same problems and cases at the same time. The GA process simplifies inter-team comparison. Lack of adequate preparation is obvious to the instructor and other students. Well-prepared groups with collaborative members generally tend to produce well-articulated, logically sound answers to GA exercises. On the other hand, work produced by ill-prepared groups and groups with uncooperative members is generally of a lesser quality. Since GA exercises are scored and included in the overall course grade, teams are motivated to work together to develop, produce, and present sound arguments (Michaelsen & Sweet, 2008).

Social Loafing

Social loafing occurs when individuals working as part of a team withhold effort because they expect for the team to compensate for their lack of effort (Frash, Kline, &

Stahura, 2004; Tucker & Brewster, 2015). As group size increases, personal input in a collaborative task progressively reduces (Frash et al., 2004; North, Linley, & Hargreaves, 2002). This is problematic for both instructors and students in TBL because the majority of the course is centered around team assignments and exercises (Chapman, Arenson, Carrigan, & Gryckiewicz, 1993; Frash et al., 2004; Robbins, 2000).

In order to control for social loafing in TBL, students need incentives to produce effective work and equity problems need to be addressed. One such control for social loafing is accountability (Michaelsen & Sweet, 2008). The literature suggests that evaluation-potential—the identifiableness of individual performance—negatively relates to social loafing. Evaluation-potential is based on the assumption that an individual is not motivated to work without credit or accountability. Thus, one's propensity to socially loaf is in direct proportion to how easily one's contribution to the work can be identified. If a team member's contribution can be identified, that team member is less likely to loaf (Frash et al., 2004).

As a tenet of TBL, accountability is built into the TBL system at both the individual and team levels. Students and teams are accountable for the quality and quantity of their work. In TBL, students are accountable to their teammates for RAP preparation and contributions to team discussions during GA exercises (Michaelsen & Sweet, 2008). This pressures students to stay current with the material (Parmelee & Hudes, 2012). Teams are accountable to the instructor for presenting answers to GA questions in a clear, logical manner using evidence from course materials to support their arguments (Michaelsen & Sweet, 2008).

Another control for social loafing in TBL is the grading system. An effective and fair grading system provides rewards for individual contributions to team exercises and effective teamwork. The grading system also needs to alleviate students' concerns about scoring group work to ensure that hardworking team members will not be shortchanged (Burgess et al., 2014).

Grading

Grades in TBL are based on a combination of individual performance, group performance, and peer evaluation (Cestone, Levine, & Lane, 2008; Michaelsen, 1992). Each component of the overall grade should be given enough weight to prove to the students that the instructor values it. The instructor needs to be personally comfortable with administering the determined weighted grading system. Furthermore, students need to feel that the grading system is reasonable and fair (Cestone et al., 2008; Michaelsen, 1992; Roebuck, 1998). TBL grades are different from grades in most group learning settings because high performers do not have to do all the work nor fear that poor performers will drag down their scores. Every team member is held accountable for their individual work and their contribution to the team (Parmelee et al., 2012).

In a TBL classroom, individual performance and team performance are both components of the final grade. Including the individual's own performance (i.e., iRAT scores and final exam grades) ensures that individuals will work to personally prepare for TBL classes (Michaelsen & Sweet, 2008). Including the team's performance (i.e., tRAT scores and GA scores) justifies the amount of effort students put into group work, supports the development of group cohesiveness, and ensures student contribution to team discussions (Michaelsen, 1992; Michaelsen & Sweet, 2008; Roebuck, 1998).

An individual's team performance score is weighted by their contribution to the team's performance, which is often based on peer evaluations. Peer evaluations are utilized because team members generally have more information to accurately evaluate students, compared with the amount of information instructors have about an individual's team contribution (Michaelsen & Sweet, 2008; Parmelee et al., 2012). Peer evaluations are widely used in TBL to ensure that students all contribute to team exercises and increase the identifiability of individual effort within the team (Frash et al., 2004). Students who do not adequately prepare for the RAP, do not actively participate in team discussions, or are often absent from class are in danger of receiving poor evaluations that negatively affect their grades. Thus, students are motivated to attend class, contribute to team discussions, participate in GA exercises, and complete RAP readings, since each individual is rewarded for his or her effort through peer evaluations (Michaelsen, 1992; Michaelsen & Sweet, 2008; Roebuck, 1998).

While TBL classes always include weighted components that factor into a final grade, the specific weights of the different components vary from classroom to classroom. Michaelsen and Sweet (2008) suggested that instructors allow the students in a class to determine the weights for team performance, individual performance, and member contribution to the team, within the parameters set by the instructor (Michaelsen & Sweet, 2008; Parmelee et al., 2012). To do this, instructors can use the exercise "Setting Grade Weights" found in Michaelsen, Cragin & Watson (1981) (Michaelsen, 1992; Roebuck, 1998). Grade weights can function as an incentive for students to motivate students to develop normative behaviors that are conducive to success in TBL, such as individual preparation, open team communication, respectful disagreement, and problem solving (Haidet et al., 2012).

Methodological Issues

Thus far, the reporting of TBL research has been rather inconsistent, making it more difficult for general conclusions to be drawn about TBL (Haidet et al., 2012). In many studies, the TBL programs are inadequately described, making it difficult for readers to distinguish the degree to which TBL was implemented (Burgess et al., 2014; Fatmi et al., 2013). Reports of TBL do not always include comparison groups or employ experimental design (Haidet et al., 2012). TBL programs also range in length, as some instructors used the method for only a few class sessions in a term, while others used TBL for the duration of a course. In some cases, only facets of TBL were adopted for use in modified TBL models (Burgess et al., 2014). Furthermore, the types of data collected as outcome measurements are inconsistent, including student satisfaction ratings, qualitative student feedback, academic outcomes (e.g., exam scores and grades), informal class observations, and attendance (Haidet et al., 2012). Sometimes, studies of TBL used concurrent cohorts and randomly assigned students to TBL or control groups. However, studies more commonly employ retrospective cohorts taught using some other pedagogical method as comparison groups. Thus, considerable heterogeneity exists in the realm of TBL research across intervention type, comparison group, and study design (Fatmi et al., 2013).

The Current Study

To date, there has not been a meta-analysis of TBL research. While TBL has seen an increase in popularity in the last few decades across a variety of fields, the question remains: How effective is TBL? While some have argued that TBL is effective (e.g., Tucker & Brewster, 2015), these arguments are usually based on single studies, which can produce a biased understanding of the effects of TBL (Bushman & Wells, 2001). Some researchers have conducted qualitative literature reviews about TBL (e.g., Burgess et al., 2014; Haidet et al., 2014), but these suffer from some of the same problems as single studies, which is why there has recently been a call for a more quantitative review of the literature (Fatmi et al., 2013).

A better approach to analyze the effectiveness of TBL is to synthesize the TBL studies quantitatively using a meta-analysis. Meta-analyses have the advantages of providing a systematic method for organizing a large number of research studies and statistically synthesizing the results (Field & Gillett, 2010). Moreover, because a meta-analysis quantifies effects across multiple studies, it can provide an indication of both the average and variability of the population effects. Measuring this variability is important, as it allows for an exploration of why effects may differ across studies (i.e., moderation). Such information could be helpful for future educators who want to utilize TBL and would supplement existing TBL research.

Hypotheses

The independent variable in this meta-analysis is whether TBL was implemented in a course or not. The dependent variable is the result of TBL administration, or lack thereof. This is measured in the form of an academic outcome. The three academic outcome types that were implemented in TBL studies were standardized examination scores, course examination scores, and final course grades.

The hypothesis of this study is that TBL will have some effect on student academic outcomes. An additional hypothesis is that country of origin (United States or other countries), outcome measure (standardized examination scores, course examination scores, or final course grades), educational level, and course subject would be moderators for TBL effectiveness.

These hypotheses are relevant to current TBL practitioners and to instructors who are interested in using TBL in a secondary or post-secondary educational setting.

CHAPTER THREE

Methods

Study Selection

I used the following search terms to find publications reporting effects of TBL: post-secondary; secondary; education; cooperative; collaborative; learn; team; team-based. It is important to use multiple databases to find studies for a meta-analysis (Wu, Aylward, Roberts, & Evans, 2012). Therefore, the following sources were searched: Academic Search Complete, PsychInfo, JSTOR, and Google Scholar. In addition, I searched the reference lists of all identified articles to find additional TBL manuscripts and obtained a bibliography compiled by the Team-Based Learning Collaborative.

Selection Criteria

To be included in the meta-analysis, studies had to meet the following criteria: (a) published before 2016; (b) refer to Michaelsen in the reference list; (c) the TBL design included student teams and required pre-class preparation, RAP, and GA exercises; (d) included academic performance as an outcome variable; (e) reported sample size, group means, and group standard deviations (or the authors must be able to provide them upon request); (f) study participants were students in secondary, post-secondary, or graduate/professional education; and (g) included data from a comparison group. The comparison group data could be collected at different time points (e.g., compare traditional teaching in semester one with TBL in semester two), or concurrently (e.g., compare traditional teaching for section one of a course with TBL in section two of a course during the same semester).

Variable Coding

One individual completed all coding procedures. For all studies, I recoded the study's reference information (e.g., authors, year, publication source), reported statistics (e.g., sample size, means, standard deviations), and data on four potential moderators: (a) country; (b) outcome measure, (c) educational level of students, and (d) course content. In addition, as indicators of study quality, I recorded: (a) research design (e.g., random assignment, cross-over study, waitlist control study, non-randomized control trial); (b) cohort status (e.g., concurrent cohort, retrospective cohort); (c) comparison pedagogical method (e.g., active learning, small group, lecture); and (d) whether I had to estimate any values from the study (e.g., mean and standard deviation were reported in bar plots). For all studies, I coded the data so that a positive mean difference indicates that TBL group was more effective than the comparison.

Moderators

For country of origin, I coded whether or not the study’s sample came from the United States. For outcome measures, I coded whether or not the academic measure was instructor-determined (i.e., course examinations, course grades) or standardized (i.e., standardized examinations). For educational level, I coded whether students were in high school, college, graduate school, or a professional program (e.g., medicine, pharmacy, veterinary science, physical therapy). For course content, I coded whether the courses were in medicine, pharmacy, social studies, science, business, law, psychology, education, veterinary science, or physical therapy.

Effect Size Measures

An *effect size* (ES) is a quantitative reflection of the magnitude of some phenomenon that is used for the purposes of addressing a question of interest (Kelley & Preacher, 2012). For this meta-analysis, I converted all the studies' results in Hedges' (1981) *g* ES. Hedges' *g* is a standardized measure of group differences that corrects the traditional *d* effect size for the bias inherent when using small sample sizes (Borenstein, 2009). In this analysis, all *g* values were calculated using the equation for independent samples because none of the studies provided correlation coefficients between Time 1 and Time 2, which is needed to calculate the dependent-sample version of the effect size.

According to Borenstein (2009), the standardized mean difference (*d*) from studies using two independent groups is calculated as:

$$d = \frac{\bar{Y}_1 - \bar{Y}_2}{S_{Within}} \tag{1}$$

where $\bar{Y}_1$ and $\bar{Y}_2$ are group sample means and S_{Within} is the within-groups standard deviation, pooled across groups. S_{Within} is calculated as:

$$S_{Within} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \tag{2}$$

where n_1 and n_2 are the group sample sizes, and S_1 and S_2 are the group standard deviations. The sampling variance of *d* is:

$$v_d = \frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2)} \tag{3}$$

which makes the standard error of *d*:

$$SE_d = \sqrt{v_d} \tag{4}$$

The correction factor *J* is used to *d* to Hedges' *g*:

$$J(df) = 1 - \frac{3}{4df - 1} \tag{5}$$

Where *df* is the degrees of freedom used to estimate S_{Within} (i.e., $n_1 + n_2 - 2$ for two independent groups). With this correction, *g*, the variance of *g* and the standard error of *g* can be calculated:

$$g = J(df)d \tag{6}$$

$$v_g = [J(df)]^2 v_d \tag{7}$$

$$SE_g = \sqrt{v_g} \tag{8}$$

Some studies did not report one or more of the values needed to calculate *d/g*. For these studies, I attempted to estimate the means, standard deviations, or sample sizes were estimated from other information in the studies (e.g., graphs, other reported statistics or ESs).

Data Analysis

Meta-Analysis Models

Arguably, the best way to analyze meta-analytic data is to use a random-effects models. (Field, 2005). Although there are varying definition of fixed and random effects (Gelman, 2005), we use the definitions typically used in the meta-analytic literature (Valentine, 2012). Conceptually, the fixed-effects model conditions on the true effects under investigation, so provides a conditional inference about the set of studies included in the meta-analysis. In other words, it assumes that the studies are estimating a single underlying population effect, and any observed variation in the observed ES across studies is due to random sampling error (i.e., v_g in Equation 7).

A random-effects model assumes that population effect sizes vary randomly from study to study. Thus, effect sizes are thought of as being sampled from a universe of possible effects, and consequently there are two measures of variability: variability within a study (i.e., sampling error) and variability between studies (Hedges & Vevea, 1998).

Random-effects models are typically better able to model effect size variability, if it exists, than fixed-effects models. Moreover, if there is little variability among studies, then the random-effects portion of the model is minimal and it becomes a fixed-effects model.

Statistically, the major difference between the two models can be seen by comparing how they calculate the weighted average ES, $\overline{ES}$. The weighted average is calculated as

$$\overline{ES} = \frac{\sum w_i ES_i}{\sum w_i} \tag{9}$$

where w_i is ith study's weight. For fixed effects models, w is

$$w = \frac{1}{v} \tag{10}$$

where v is the sampling variance of the ES (i.e., for this study, v is v_g, as defined in Equation 7.

For a random-effects model, w is calculated as

$$w = \frac{1}{\tau^2 + v} \tag{11}$$

where τ^2 is the population variability of the effects (i.e., a measure of how much variability there is in ES across studies).

When moderators are included, the model becomes a "mixed-effects" model, which combines fixed-effects and random-effects model. This can be seen in the equation for the mixed-effects model

$$ES_i = b_0 + b_1 x_{i1} + \cdots + e_i + u_i, \tag{12}$$

where u_i is the population variability (thought to follow a normal distribution with a mean of 0 and variance of τ^2), e_i is the sampling variability (thought to follow a normal distribution with a mean of 0 and variance of v_i), x_{i1} is the value of the first moderator, b_0 is the intercept (i.e., average ES when the values of all moderator variables are equal to zero), and b_1 is the average change in the ES for a one unit increase in x_{i1}.

Some studies reported more than one ES. While there are differing views on how to handle this situation, I compared models that did not (i.e., typical random effects model) and did account for the nested data. The model that accounted for the nested data was a "three-level" multilevel model that added an extra variance term for within citation variance (Konstantopoulos, 2011).

Data Analysis Steps for Study

For the current study, I analyzed the data using the following steps:

1. Calculate Hedges' *g* for each study that met the inclusion criteria.
2. Estimate the average effect size for all studies, as well as heterogeneity in the effect sizes. To estimate this heterogeneity, I used two methods: Q and I^2 statistics.

The Q statistic is calculated as

$$Q = \sum w_i \,(ES_i - \overline{ES})^2. \tag{13}$$

Since Q is a weighted sum of squares, it is approximately distributed as a χ^2 with k - 1 degrees of freedom (*df*), where k is the number of studies. Thus, not only can it be used to test heterogeneity, but Q statistics can be compared between nested models.

The I^2 measure of this inconsistency in the ESs. It is a variation of the Q statistic, but provides the variability in ESs due to variability in the population of studies rather than sampling error. It is calculated as

$$I^2 = \left(\frac{Q - k - 1}{Q}\right) \times 100 \tag{14}$$

3. Compare models that do, and do not, account for studies being nested in citations.
4. Examine if there were any moderator effects by including each of the moderator variables in the model, singly, and determining if it explained part of the heterogeneity in effect sizes. I determined this my examining Akieke's (1974) information criterion (AIC) values across models. AIC is a measure of model fit that balances explain the data with model parsimony. The AIC metric itself is difficult to interpret, but when comparing models smaller values indicate better models.
5. Examine if there was any publication bias (i.e., that statistically significant findings were more likely to be published than non-significant findings; Dickersin, 2005). I did this by using a funnel plot (Light & Pillemer, 1984;

Sterne, Becker, & Egger, 2005). A funnel plot is a scatter plot of effect sizes against a measure of the study's size (e.g., standard errors), which has shown to be useful in examining publication bias (Egger, Jüni, Bartlett, Holenstein, & Sterne, 2003). If there is bias, then the funnel plot will appear asymmetric. In a sample without publication bias, the plot will tend to have the shape of a funnel. In the case of outliers, the data associated with the outliers were checked to ensure that the outliers were not caused by data entry errors.

For the data analysis, I used the R statistical program (R Core Team, 2014). I calculated all effect sizes, estimated all meta-analytic models, and created funnel plots using the *metafor* package (Viechtbauer, 2010).

CHAPTER FOUR

Results

Initially, I found a total of 462 individual citations using the search terms. For each citation, I read the abstracts and reference lists of all the studies to see if the articles were about TBL. 200 articles were found to be relevant to TBL. 162 studies were about TBL but were excluded from the meta-analysis because the studies did not meet some aspect of the inclusion criteria. The 162 studies included the RAP and GA processes and cited Michaelsen in the reference section, but were omitted from the meta-analysis because not enough data was reported, there was no comparison group in the study, modified TBL was used, or the reported outcome measure was not an academic one (e.g., student engagement, attendance rate, satisfaction rating). 38 citations fulfilled the inclusion criteria and were included in the meta-analysis. The 38 citations produced 146 studies (see Appendix A for citations, studies, and effect sizes).

Estimation of certain values was required for some of the cases in this study. Five of the articles reported statistics using histograms in lieu of reporting exact statistics, which led to the estimation of means and standard deviation values. The sample size for the comparison group in one study was imputed from the sample size of the experimental group. The g value for one study was manually calculated and added to the analysis. In one study, the means and standard deviations of both the TBL and control groups were estimated from the reported medians and inter-quartile ranges.

Overview

Accounting for Nested Data

First, I fit an unconditional (baseline) model, which has no predictors and accounted for the nested nature of the data. This provides the weighted mean for the entire sample. The results are in Table 1. Next I fit a typical random effects model to examine the effect of not accounting for the nested nature of the data. The results are also shown in Table 1. The difference between models is minimal. In both models, average ES for TBL is positive and indicates that, on average, the TBL groups' performance on the outcomes was approximately 1/3 of a standard deviation higher than the comparison groups. Likewise, the AIC values for both models were very similar, although the model that does not account for the nested nature of the studies has a slightly smaller value. Consequently, I used the typical random effects model baseline model for the rest of the analyses.

Table 1
Results from Meta-Analytic Models

Model	AIC	b	SE b
Baseline-3 level	275.40	.32	.06
Baseline-typical	274.61	.35	.05

Moderator Models

Table 2

Results from Random Effects Meta-Analytic Models

Model	*k*	AIC	b	SE b	τ^2 (SE τ^2)	*Q* (*df*)	I^2
Baseline	146	274.61	.35	.05	.35 (.04)	3142.67 (145)	95.45
Country of Origin		274.92			.35 (.05)	3128.29 (144)	94.44
United States	132		.34	.05			
Foreign	14		.49	.17			
Outcome type		273.90			.35 (.05)	3072.59 (143)	95.41
Standardized exam	53		.26	.09			
Course exam	75		.17	.11			
Course grade	18		-.02	.17			
Education level		272.68			.35 (.04)	2980.28 (141)	95.35
Secondary	24		.19	.12			
Undergraduate	16		.19	.20			
Medical Doctor	65		.29	.15			
Doctor of Pharmacy	30		.04	.17			
Other Graduate Levels	11		.04	.23			
Course Subject		273.69			.35 (.05)	2972.80 (141)	95.38
Medicine	74		.46	.07			
Pharmacy	30		-.24	.13			
Social Studies	21		-.26	.16			
Science	10		-.17	.21			
Other Subjects	11		-.20	.20			

The results from the baseline model indicated that there was substantial variability in the data not due to sampling error (see Table 2). Not only was the Q statistic large and statistically significant, but I^2 was 95.45, indicating that approximately 95% of the variability in ESs is due to variability in the population of studies rather than sampling error. Consequently, I next examined if any moderator could help explain this variability.

First, I added country of origin. Only 14 of the articles that fulfilled the inclusion criteria were from outside the United States. Thus, it was not feasible to conduct moderator analyses on each of the countries represented in the data. Instead, all of the articles from outside the United States were condensed into a single group. Thus, the two levels of the country of origin moderator were: United States, and non-United States.

The analysis of this moderator showed that whether TBL was conducted in the United States or in other countries explained a minimal amount of variability in the data. The difference in Q statistic values from the baseline model was not significant (Q= .81, df=1, p=.37) and the I^2 value indicated that 95% of the variability in ESs is still due to variability in the population of studies Moreover, the AIC value was higher for the model with the moderator than the baseline model. Thus, it appears that country of origin is not a viable moderator of TBL's effect.

The second moderator was outcome measure. There were three types of academic outcome measures in the TBL studies: (a) standardized exam score, (b) course exam score, and (c) course grade. The analysis of this moderator showed that outcome type explained a minimal amount of variability in the data. Although the AIC value was slightly smaller for the moderator model than the baseline model, the reduction of approximately one AIC unit is small. Moreover, the difference in Q statistic values from the baseline model was not significant (Q= 2.97, df=1, p=..23) and the I^2 value indicated that 95% of the variability in ESs is still due to variability in the population of studies Consequently, it appears that the outcome measure used in a study is not a viable moderator of TBL's effect.

The third moderator included in this analysis was educational level. TBL was implemented in courses with students ranging from middle school to continuing professional education. As a large age range is represented in these studies, this moderator was included to see if the level of education would account for any variability in the effects of TBL. Eight educational levels were originally coded: (a) secondary, (b) undergraduate, (c) medical doctor, (d) doctor of pharmacy, (e) graduate, (f) post-graduate, (g) veterinary physician, and (h) doctor of physical therapy. However, there were few articles in the graduate, post-graduate, veterinary physician, and doctor of physical therapy categories. Thus, these four levels of education were collapsed into a single level of “other graduate education”, bringing the total number of levels to five.

The analysis of this moderator showed that education level explained a minimal amount of variability in the data. Although the AIC value was slightly smaller for the moderator model than the baseline model, the reduction of approximately two AIC units is small. Moreover, the difference in Q statistic values from the baseline model was not significant (Q= 6.44, df=4, p=.17) and the I^2 value indicated that 95% of the variability in ESs is still due to variability in the population of studies Consequently, it appears that education level is not a viable moderator of TBL's effect.

The last moderator examined in the analysis was course subject. There is a wide variety of subjects that are taught using TBL. By using course subject as a moderator, it is possible to determine whether all subjects can be taught equally effectively using TBL, or if certain subjects are more suited for TBL implementation than others. Ten course subjects were originally coded: medicine, pharmacy, social studies, science, business, law, psychology, education, veterinary science, and physical therapy. However, there were few articles in the last six categories in the analysis. Thus, business, law, psychology, education, veterinary science and physical therapy were collapsed into a single level of “other subjects”, bringing the total number of levels to five.

Although the AIC value was slightly smaller for the moderator model than the baseline model, the reduction of approximately one AIC unit is small. Moreover, the difference in Q statistic values from the baseline model was not significant (Q= 5.44, df=4, p=.25) and the I^2 value indicated that 95% of the variability in ESs is still due to variability in the population of studies Consequently, it appears that course subject is not a viable moderator of TBL's effect.

Publication Bias

To examine publication bias, I created a funnel plot of the ESs (Figure 4). In general, the funnel plot is symmetric. Although there is some asymmetry at the very bottom with a gap appears in the bottom left of the plot. This could indicate that there some bias in that studies with smaller sample sizes (i.e., higher sampling variance) and lower or negative effects are less likely to be published.

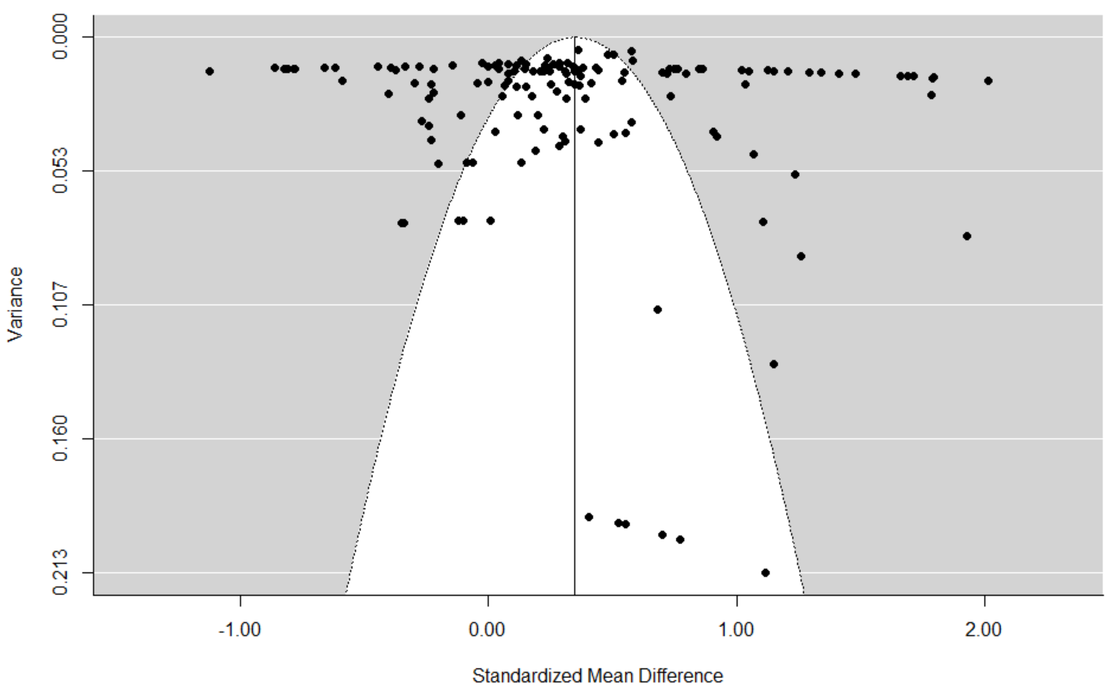

Figure 4. Funnel plot with variance and standardized mean difference.

To examine if studies are possibly missing, I "trimmed and filled" the funnel plot (Duval & Tweedie, 2000a, b). The trim and fill method is a rank-based data augmentation technique to estimate the number of studies missing from a meta-analysis. The augmented funnel plot is shown in Figure 5. The plot is virtually identical to the one in Figure 4, indicating that indicates that if there is a publication bias in the TBL studies, it is small and likely has minimal influence on the over results of this study.

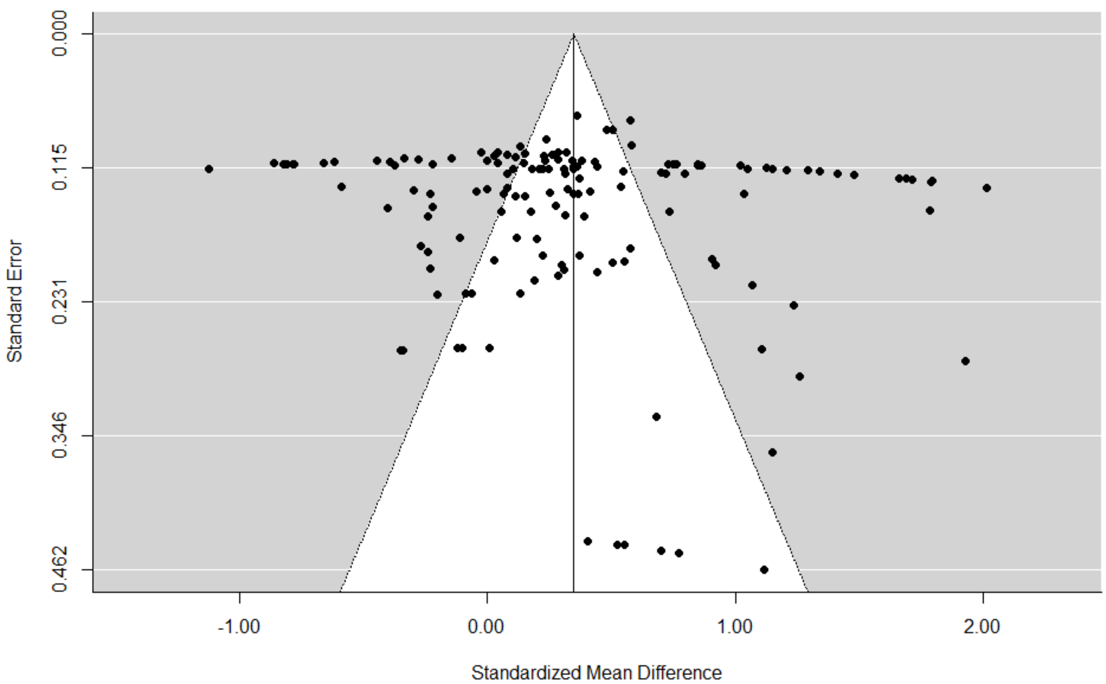

Figure 5. Funnel plot with standardized error and standardized mean difference.

CHAPTER FIVE

Discussion

Major Findings

Overall, there was a positive effect in the baseline model. TBL seems to produce better academic outcomes than the comparison pedagogical methods, on average across all studies. After examining the four possible moderators of TBL effectiveness, two of the factors accounted for some of the variability within the data (country of origin and outcome measure) and two of the factors did not account for the variability within the data (educational level and course subject). Since the moderators could not explain much of the variability in effect sizes, the moderator variables were not analyzed in combination with one another. Thus, the emphasis of this study should be on the baseline model.

Based on the results of this analysis, it appears that TBL is a method that deserves more investigation as pedagogical technique. To date, little in the education effectiveness literature has examined TBL---e.g., Hattie (2012) does not even mention it in his mega-analysis of learning techniques. TBL might especially be useful for courses with material that students can reasonably be expected to grasp outside of class or that emphasize practical application of course knowledge. By using TBL, instructors can reduce the amount of time spent on lectures, while ensuring that students study course material independently. Then, instructors can spend more time on application exercises to provide students with opportunities to utilize content knowledge in real-life scenarios.

Limitations

The results from this meta-analysis need to interpreted with some caution due to the study sampling and research design used in many TBL studies. Out of the 200 articles about TBL found in the article search, less than a fifth of the articles (38) could be used in the meta-analysis. In many cases, basic descriptive statistics (i.e., means, standard deviations, and sample sizes) were simply not reported. In other cases, the outcome measures were not of an academic nature. Instead of reporting examination scores or course grades, satisfaction ratings, attendance rates, levels of student engagement and other non-academic outcome measures were used. In many cases, TBL studies failed to include a comparison group. Sometimes a pre/post design was used, resulting in difficulties with interpretation because it is unclear how much of the change in knowledge is due to the TBL methodology and how much of the change in knowledge is due to the fact that the students were enrolled in a course at all.

Of the articles that were selected for the meta-analysis, there was a wide range in research designs. While some studies employed randomization and concurrent cohorts, most of the articles displayed a lack of randomization and the use of retrospective cohorts. In the cases of non-randomization and retrospective cohorts, it is often difficult to determine whether or not the student samples in the experimental and control groups were actually comparable.

Furthermore, many of the TBL studies were often pilot studies explaining the results of pioneering TBL usage in a field that previously did not implement TBL. These first-time administrations may have some issues or problems that have yet to be worked out by the instructors. This phenomenon raises several questions: is it fair to compare the academic outcomes of one cohort of students with an experienced lecturer to the academic outcomes of anothercohort of students with an inexperienced TBL instructor? Would TBL studies hold more weight if the instructors had a few years of experience before TBL courses are compared—concurrently or retrospectively—with courses taught using other pedagogical methods?

Other Limitations

In this meta-analysis, there were far fewer TBL administrations outside of the United States than within the United States. This did not indicate that TBL was not implemented in other countries. Rather, many of the studies conducted outside of the United States simply did not meet the inclusion criteria, limiting the scope of the meta-analysis to a less internationally representative sample.

Not all educational levels or academic subjects were represented equally in the analysis, as more studies in medical and pharmacy programs were included in the analysis than studies in other fields.. According to the literature, TBL is implemented in many educational settings and covers a wide variety of subjects. However, this meta-analysis is not reflective of such a claim. There needs to be more—and better—reporting of TBL research conducted in non-health fields, so that a broader and more representative analysis of TBL research can be conducted in the future.

Lastly, several studies reported statistics that required some estimation by the data coder. Five studies (Ingram & Adams, 2003; Johnson et al., 2014; Anwar, Shaikh, Dash, & Khurshid, 2012; Nieder, Parmelee, Stolfi, & Hudes, 2005; Kenny, McLaren, Blissenden, & Villios, 2015) did not report means and standard deviations, but reported data using histograms. The coder calculated the means and standard deviations of these studies by estimating the values represented in the histograms and performing calculations using frequency tables. In one study (Nyindo et al., 2014) the comparison group sample size was not reported, so the value was imputed based on the TBL group sample size. One study did not report means, standard deviations, or sample sizes (Nicoll-Senft, 2009), but the g value and g variance were calculated by hand, based on the F statistic reported in the results section. In one study, the means and standard deviations were estimated based on the reported medians and interquartile range values (Punja, Kaludi, Pai, Rao, & Dhar, 2014).

Theoretical Implications

From the results of this meta-analysis, TBL seems to be effective. This supports the information found in the literature that claims TBL to be an effective pedagogical method. The results of this study follow the general trend set by previous systematic research reviews (Burgess, McGregor, & Mellis, 2014; Fatmi et al., 2013; Haidet, Kubitz, & McCormack, 2014). TBL practitioners can use the conclusions of this meta-analysis to support their own TBL implementation.

The findings of this meta-analysis can be tentatively generalized to student populations in the health field (especially to student populations in medical and pharmacology schools) as well as secondary social studies education, since these populations were represented to a greater extent in the analysis. While the findings in this analysis may be extended to contexts outside of education in social studies and the health field, the effectiveness of TBL in other fields needs to be evaluated with future research, to determine the extent to which the generalizations of these findings can be made.

Methodological Implications

Those who are interested in conducting future research in TBL can benefit from consulting resources about best practices for conducting and reporting TBL research. There is a need for more robust primary research to be performed in TBL, complete with thorough and descriptive reporting (Fatmi et al., 2013).

Perspective: Guidelines for Reporting Team-Based Learning Activities in the Medical and Health Sciences Education Literature by Haidet et al. (2012) is a good resource to follow for information about standardized reporting of TBL findings. The reporting of TBL research thus far has been rather inconsistent, making it more difficult for general conclusions to be drawn about TBL. In this set of guidelines, seven core elements of TBL were identified: (a) team formation; (b) readiness assurance (RA); (c) immediate feedback; (d) sequencing of in-class problem solving; (e) the "Four S's"; (f) incentive structure; (g) peer review (Haidet et al., 2012).

Team-Based Learning: A Practical Guide: AMEE Guide No. 65 by Parmelee, Michaelsen, Cook & Hudes (2012) is a valuable resource to educators who want to implement TBL in their own classrooms, but do not know where to start. It includes a clear, detailed breakdown of the TBL process from the perspectives of both students and teachers and provides steps for developing an effective TBL course (Parmelee et al., 2012)

If the recommendations in these guides are followed, a future meta-analysis on TBL would be able to more thoroughly account for overall effects of TBL, as well as the moderator effects of TBL. A future study would be able to determine:

1. Whether country of origin still has an effect on variability. If so, a more detailed breakdown of the country of origin might be possible. Instead of only analyzing

the effects of TBL inside and outside the United States, it might be possible for future studies to analyze TBL implementation based on a more representative comparison (i.e. a comparison between geographical regions, cultures, or continents).

2. Whether outcome measure type still has an effect on variability.
3. Whether educational level has a meaningful effect on variability.
4. Whether course subject has a meaningful effect on variability.
5. Whether the type of comparison group used in the study (e.g., small group, active learning, lecture) has a meaningful effect on variability.
6. Whether the type of cohort in a study (e.g., retrospective cohort versus concurrent cohort) has a meaningful effect on variability.

APPENDICES

APPENDIX A

Table 3

Studies Included in the Meta-Analysis

Author (Year)	TBL	TBL N	Comparison	Comp N	*g*	*g Var*	Course Subject	Outcome Measure	Country	Education Level
Anwar, Shaikh, Dash, & Khurshid (2012)*	Exam Score	104	Exam Score	113	*0.25*	*0.02*	Pathology	Course Exam	United Arab Emirates	Medical School
Bleske et al. (2014)	Multiple Choice Recall Score	97	Multiple Choice Recall Score	85	*-0.40*	*0.02*	Therapeutics	Course Exam	United States	Pharmacy School
Bleske et al. (2014)	Multiple Choice Application Score	97	Multiple Choice Application Score	85	*-0.22*	*0.02*	Therapeutics	Course Exam	United States	Pharmacy School
Bleske et al. (2014)	Essay Score	97	Essay Score	85	*-0.22*	*0.02*	Therapeutics	Course Exam	United States	Pharmacy School
Carmichael (2009)	Data Interpretatio n Score	108	Data Interpretation Score	107	*0.35*	*0.02*	Introduction to Biology	Course Exam	United States	Undergraduate
Dinan & Frydrychowski (1995)	Final Exam Score 1993	36	Final Exam Score (1990, 1991, 1992)	106	*0.55*	*0.04*	Organic Chemistry	Course Exam	United States	Undergraduate
Goldberg & Dintzis (2007)	Final Exam (2006)	121	Final Exam (2005)	120	*1.03*	*0.02*	Physiology and Histology	Course Exam	United States	Medical School

Author (Year)	TBL	TBL N	Comparison	Comp N	*g*	*g Var*	Course Subject	Outcome Measure	Country	Education Level
Haberyan (2007)	Final Grade	40	Final Grade	48	*1.24*	*0.05*	Industrial/Organizational Psychology	Course Grade	United States	Undergraduate
Huggins & Stamatel (2015)	Short Answer Exam Score	101	Short Answer Exam Score	74	*0.18*	*0.02*	Sociology	Course Exam	United States	Undergraduate
Huggins & Stamatel (2015)	Essay Exam Score	101	Essay Exam Score	74	*0.06*	*0.02*	Sociology	Course Exam	United States	Undergraduate
Huitt, Killins, & Brooks (2015)	Final Exam Score (2012)	88	Final Exam Score (2010-2011)	124	*0.12*	*0.02*	Gross Anatomy Lab	Course Exam	United States	Physical Therapy School
Huitt, Killins, & Brooks (2015)	Course Grade (2012)	88	Course Grade (2010-2011)	124	*0.15*	*0.02*	Gross Anatomy Lab	Course Exam	United States	Physical Therapy School
Ingram & Adams (2003)*	Course Grade (1996-1998)	120	Course Grade (1994-1996)	219	*0.35*	*0.01*	Finance	Course Grade	United States	Undergraduate
Jarjoura, Tayeh, & Zgheib (2015)	Course Grade	60	Course Grade	30	*0.13*	*0.05*	Biology	Course Exam	Lebanon	7th Grade
Johnson et al. (2014)*	Course 1 Grade (Year 3)	114	Course 1 Grade (Year 2)	103	*-0.04*	*0.02*	Pharmacotherapeutics	Course Grade	United States	Pharmacy School
Johnson et al. (2014)*	Course 1 Grade (Year 4)	110	Course 1 Grade (Year 2)	103	*-0.23*	*0.02*	Pharmacotherapeutics	Course Grade	United States	Pharmacy School
Johnson et al. (2014)*	Course 1 Grade (Year 5)	125	Course 1 Grade (Year 2)	103	*0.00*	*0.02*	Pharmacotherapeutics	Course Grade	United States	Pharmacy School
Johnson et al. (2014)*	Course 1 Grade (Year 6)	128	Course 1 Grade (Year 2)	103	*0.08*	*0.02*	Pharmacotherapeutics	Course Grade	United States	Pharmacy School

(Continued)

Author (Year)	TBL	TBL N	Comparison	Comp N	*g*	*g Var*	Course Subject	Outcome Measure	Country	Education Level
Johnson et al. (2014)*	Course 2 Grade (Year 3)	113	Course 2 Grade (Year 2)	102	*0.37*	*0.02*	Pharmacotherapeu tics	Course Grade	United States	Pharmacy School
Johnson et al. (2014)*	Course 2 Grade (Year 4)	109	Course 2 Grade (Year 2)	102	*0.07*	*0.02*	Pharmacotherapeu tics	Course Grade	United States	Pharmacy School
Johnson et al. (2014)*	Course 2 Grade (Year 5)	123	Course 2 Grade (Year 2)	102	*0.41*	*0.02*	Pharmacotherapeu tics	Course Grade	United States	Pharmacy School
Johnson et al. (2014)*	Course 2 Grade (Year 6)	130	Course 2 Grade (Year 2)	102	*0.32*	*0.02*	Pharmacotherapeu tics	Course Grade	United States	Pharmacy School
Kenny, McLaren, Blissenden, & Villios (2015)*	Final Exam Score (2013)	93	Final Exam Score (2009)	171	*-0.59*	*0.02*	Taxation Law	Course Exam	Australia	Undergraduate
Kenny, McLaren, Blissenden, & Villios (2015)*	Final Exam Score (2014)	82	Final Exam Score (2009)	171	*-0.30*	*0.02*	Taxation Law	Course Exam	Australia	Undergraduate
Kent, Wanzek, Swanson, & Vaughn (2015)	Vocab Items	16	Vocab Items	8	*1.11*	*0.21*	U. S. History	Stand. Exam	United States	11th Grade
Kent, Wanzek, Swanson, & Vaughn (2015)	Comprehens ion Items	16	Comprehensi on Items	8	*0.52*	*0.19*	U. S. History	Stand. Exam	United States	11th Grade
Kent, Wanzek, Swanson, & Vaughn (2015)	Total Score	16	Total Score	8	*0.70*	*0.20*	U. S. History	Stand. Exam	United States	11th Grade

(Continued)

Author (Year)	TBL	TBL N	Comparison	Comp N	*g*	*g Var*	Course Subject	Outcome Measure	Country	Education Level
Koles, Nelson, Stolfi, Parmelee, & DeStephen (2005)	Immune, Neoplastic, Cardiovascular, Parathyroid Exam Score	41	Immune, Neoplastic, Cardiovascular, Parathyroid Exam Score	39	*-0.06*	*0.05*	Pathology	Course Exam	United States	Medical School
Koles, Nelson, Stolfi, Parmelee, & DeStephen (2005)	Genetic, Muscle, Breast, Liver Exam Score	39	Genetic, Muscle, Breast, Liver Exam Score	41	*-0.20*	*0.05*	Pathology	Course Exam	United States	Medical School
Levine et al. (2004)	Licensing Exam Subject Test Score (2005)	133	Licensing Exam Subject Test Score (2004)	130	*0.37*	*0.02*	Psychiatry Clerkship	Stand. Exam	United States	Medical School
Levine et al. (2004)	Licensing Exam Subject Test Score (2005)	133	Licensing Exam Subject Test Score (2003)	147	*0.31*	*0.01*	Psychiatry Clerkship	Stand. Exam	United States	Medical School
Malone & Spieth (2012)	Urinary Surgery Test Score	38	Urinary Surgery Test Score	54	*0.19*	*0.05*	Urinary Systems Disorders	Course Exam	United States	Veterinary School
Malone & Spieth (2012)	Urinary Surgery Repeat Exam Score	19	Urinary Surgery Repeat Exam Score	20	*0.68*	*0.11*	Urinary Systems Disorders	Course Exam	United States	Veterinary School
McInerney & Fink (2003)	Final Exam Score (2001)	63	Final Exam Score (2000)	60	*0.58*	*0.03*	Microbial Physiology	Course Exam	United States	Undergraduate
McInerney & Fink (2003)	Final Exam Score (2002)	58	Final Exam Score (2000)	60	*0.91*	*0.04*	Microbial Physiology	Course Exam	United States	Undergraduate
Metoyer, Miller, Mount, & Westmoreland (2014)	Final Exam Score (2011)	95	Final Exam Score (2010)	87	*0.73*	*0.02*	Physics	Course Exam	United States	Undergraduate

(Continued)

Author (Year)	TBL	TBL N	Comparison	Comp N	*g*	*g Var*	Course Subject	Outcome Measure	Country	Education Level
Mody, Kiley, Gawron, Garcia, & Hammond (2013)	Course Exam	69	Course Exam	61	*0.20*	*0.03*	Family Planning	Course Exam	United States	Medical School
Nicoll-Senft (2009)*	Quiz Score	NA	Quiz Score	NA	*1.15*	*0.13*	Special Education	Course Exam	United States	Graduate (MA in Special Education)
Nieder, Parmelee, Stolfi, & Hudes (2005)*	Exam Score (2002)	95	Exam Score (1999-2001)	91	*0.27*	*0.02*	Anatomy	Course Grade	United States	Medical School
Nyindo et al. (2014)*	Ectoparasite Final Exam Score (2012)	37	Ectoparasite Final Exam Score (2011)	37	*1.93*	*0.08*	Ectoparasites	Course Exam	Tanzania	Undergraduate
Nyindo et al. (2014)*	Ectoparasite Final Exam Score (2012)	121	Ectoparasite Final Exam Score (2011)	121	*1.78*	*0.02*	Ectoparasites	Course Exam	Tanzania	Medical School
Okubo et al. (2012)	Clinical Reasoning Ability Score (2005-2007)	308	Clinical Reasoning Ability Score (2008-2010)	298	*0.50*	*0.01*	Clinical Reasoning	Stand. Exam	Japan	Medical School
Okubo et al. (2012)	Clinical Examination Score (2005-2007)	308	Clinical Examination Score (2008-2010)	297	*0.48*	*0.01*	Clinical Reasoning	Stand. Exam	Japan	Medical School

(Continued)

Author (Year)	TBL	TBL N	Comparison	Comp N	*g*	*g Var*	Course Subject	Outcome Measure	Country	Education Level
Persky & Pollack (2011)	Standardized Physiology Exam Basic Score	153	Standardized Physiology Exam Basic Score	146	*0.25*	*0.01*	Physiology	Stand. Exam	United States	Pharmacy School
Persky & Pollack (2011)	Standardized Physiology Exam Applied Score	153	Standardized Physiology Exam Applied Score	146	*0.70*	*0.01*	Physiology	Stand. Exam	United States	Pharmacy School
Persky & Pollack (2011)	Standardized Physiology Exam Overall Score	153	Standardized Physiology Exam Overall Score	146	*0.55*	*0.01*	Physiology	Stand. Exam	United States	Pharmacy School
Persky (2012)	Clinical Course Examination Score	151	Clinical Course Examination Score (Lecture)	144	*0.72*	*0.01*	Foundational Pharmacokinetics	Course Exam	United States	Pharmacy School
Persky (2012)	Clinical Course Examination Score	151	Clinical Course Examination Score (Small Group)	435	*0.58*	*0.01*	Foundational Pharmacokinetics	Course Exam	United States	Pharmacy School
Punja, Kalludi, Pai, Rao, & Dhar (2014)*	Multiple Choice Score	128	Multiple Choice Score	113	*0.54*	*0.02*	Anatomy	Course Exam	India	Undergraduate
Redwanski (2012)	Final Grade (2007)	55	Final Grade (2006)	51	*0.03*	*0.04*	Drug Information Course	Course Grade	United States	Pharmacy School
Redwanski (2012)	Final Grade (2008)	60	Final Grade (2006)	51	*0.23*	*0.04*	Drug Information Course	Course Grade	United States	Pharmacy School

(Continued)

Author (Year)	TBL	TBL N	Comparison	Comp N	*g*	*g Var*	Course Subject	Outcome Measure	Country	Education Level
Tan et al. (2011)	Exam 1 Score	49	Exam 1 Score	49	*0.44*	*0.04*	Clinical Neurology (Neurological Emergencies & Neurological Localization)	Course Exam	Singapor e	Medical School
Tan et al. (2011)	Exam 2 Score	49	Exam 2 Score	49	*1.07*	*0.05*	Clinical Neurology (Neurological Emergencies & Neurological Localization)	Course Exam	Singapor e	Medical School
Thomas & Bowen (2011)	Hypertensio n, Headache, Low Back Pain Score	52	Hypertension, Headache, Low Back Pain Score	60	*0.37*	*0.04*	Ambulatory Medicine	Course Exam	United States	Medical School
Thomas & Bowen (2011)	Diabetes, Depression, Cough Score	60	Diabetes, Depression, Cough Score	52	*0.92*	*0.04*	Ambulatory Medicine	Course Exam	United States	Medical School
Vasan, DeFouw, & Holland (2008)	Head & Neck Score (2004)	168	Head & Neck Score (2003)	173	*0.86*	*0.01*	Medical Gross Anatomy & Embryology	Course Exam	United States	Medical School
Vasan, DeFouw, & Holland (2008)	Head & Neck Score (2005)	178	Head & Neck Score (2003)	173	*1.69*	*0.02*	Medical Gross Anatomy & Embryology	Course Exam	United States	Medical School
Vasan, DeFouw, & Holland (2008)	Head & Neck Score (2006)	176	Head & Neck Score (2003)	173	*1.48*	*0.01*	Medical Gross Anatomy & Embryology	Course Exam	United States	Medical School

(Continued)

Author (Year)	TBL	TBL N	Comparison	Comp N	*g*	*g Var*	Course Subject	Outcome Measure	Country	Education Level
Vasan, DeFouw, & Holland (2008)	Thorax, Back, Upper Extremity Score (2004)	168	Thorax, Back, Upper Extremity Score (2003)	173	*0.85*	*0.01*	Medical Gross Anatomy & Embryology	Course Exam	United States	Medical School
Vasan, DeFouw, & Holland (2008)	Thorax, Back, Upper Extremity Score (2005)	178	Thorax, Back, Upper Extremity Score (2003)	173	*1.29*	*0.01*	Medical Gross Anatomy & Embryology	Course Exam	United States	Medical School
Vasan, DeFouw, & Holland (2008)	Thorax, Back, Upper Extremity Score (2006)	176	Thorax, Back, Upper Extremity Score (2003)	173	*1.71*	*0.02*	Medical Gross Anatomy & Embryology	Course Exam	United States	Medical School
Vasan, DeFouw, & Holland (2008)	Abdomen, Pelvis, Perineum, Lower Extremity Score (2004)	168	Abdomen, Pelvis, Perineum, Lower Extremity Score (2003)	173	*0.38*	*0.01*	Medical Gross Anatomy & Embryology	Course Exam	United States	Medical School
Vasan, DeFouw, & Holland (2008)	Abdomen, Pelvis, Perineum, Lower Extremity Score (2005)	178	Abdomen, Pelvis, Perineum, Lower Extremity Score (2003)	173	*1.12*	*0.01*	Medical Gross Anatomy & Embryology	Course Exam	United States	Medical School
Vasan, DeFouw, & Holland (2008)	Abdomen, Pelvis, Perineum, Lower Extremity Score (2006)	176	Abdomen, Pelvis, Perineum, Lower Extremity Score (2003)	173	*1.34*	*0.01*	Medical Gross Anatomy & Embryology	Course Exam	United States	Medical School

(Continued)

Author (Year)	TBL	TBL N	Comparison	Comp N	*g*	*g Var*	Course Subject	Outcome Measure	Country	Education Level
Vasan, DeFouw, & Holland (2008)	Licensing Exam Score (2004)	168	Licensing Exam Score (2003)	173	*0.75*	*0.01*	Medical Gross Anatomy & Embryology	Stand. Exam	United States	Medical School
Vasan, DeFouw, & Holland (2008)	Licensing Exam Score (2005)	178	Licensing Exam Score (2003)	173	*1.02*	*0.01*	Medical Gross Anatomy & Embryology	Stand. Exam	United States	Medical School
Vasan, DeFouw, & Holland (2008)	Licensing Exam Score (2006)	176	Licensing Exam Score (2003)	173	*1.80*	*0.02*	Medical Gross Anatomy & Embryology	Stand. Exam	United States	Medical School
Vaughn et al. (2013)	Reading Comprehension Score	203	Reading Comprehension Score	119	*0.31*	*0.01*	Reading Comprehension	Stand. Exam	United States	8th Grade
Vaughn et al. (2013)	Knowledge Acquisition Score	218	Knowledge Acquisition Score	121	*0.36*	*0.01*	Social Studies	Stand. Exam	United States	8th Grade
Vaughn et al. (2013)	Reading Comprehension in Social Studies Score	213	Reading Comprehension in Social Studies Score	126	*0.44*	*0.01*	Social Studies Reading Comprehension	Stand. Exam	United States	8th Grade
Wamsley et al. (2013)	History Taking Score	32	History Taking Score	24	*-0.35*	*0.07*	Screening & Brief Intervention	Stand. Exam	United States	Post-Graduate (Medical Resident)
Wamsley et al. (2013)	Brief Intervention Score	32	Brief Intervention Score	24	*1.26*	*0.09*	Screening & Brief Intervention	Stand. Exam	United States	Post-Graduate (Medical Resident)
Wamsley et al. (2013)	Patient-Physician Interaction Score	32	Patient-Physician Interaction Score	24	*-0.12*	*0.07*	Screening & Brief Intervention	Stand. Exam	United States	Post-Graduate (Medical Resident)

(Continued)

Author (Year)	TBL	TBL N	Comparison	Comp N	*g*	*g Var*	Course Subject	Outcome Measure	Country	Education Level
Wamsley et al. (2013)	Substance Use Disorder Screening Score	32	Substance Use Disorder Screening Score	24	*-0.10*	*0.07*	Screening & Brief Intervention	Stand. Exam	United States	Post-Graduate (Medical Resident)
Wamsley et al. (2013)	Substance Use Disorder Assessment Score	32	Substance Use Disorder Assessment Score	24	*-0.34*	*0.07*	Screening & Brief Intervention	Stand. Exam	United States	Post-Graduate (Medical Resident)
Wamsley et al. (2013)	Overall Brief Intervention Satisfaction Score	32	Overall Brief Intervention Satisfaction Score	24	*0.01*	*0.07*	Screening & Brief Intervention	Stand. Exam	United States	Post-Graduate (Medical Resident)
Wanzek et al. (2014)	Total Score	246	Total Score	144	*0.23*	*0.01*	Social Studies	Stand. Exam	United States	11th Grade
Wanzek et al. (2014)	Reading Comprehension Score	253	Reading Comprehension Score	141	*0.03*	*0.01*	Social Studies	Stand. Exam	United States	11th Grade
Wanzek, Kent, Vaughn, Swanson, Roberts, & Haynes (2015)	Total Score	184	Total Score	144	*-0.22*	*0.01*	U. S. History	Stand. Exam	United States	8th Grade
Wanzek, Kent, Vaughn, Swanson, Roberts, & Haynes (2015)	Written Essay: Content	179	Written Essay: Content	147	*0.15*	*0.01*	U. S. History	Stand. Exam	United States	8th Grade
Wanzek, Kent, Vaughn, Swanson, Roberts, & Haynes (2015)	Written Essay: Support	179	Written Essay: Support	147	*0.04*	*0.01*	U. S. History	Stand. Exam	United States	8th Grade

Continued)

Author (Year)	TBL	TBL N	Comparison	Comp N	*g*	*g Var*	Course Subject	Outcome Measure	Country	Education Level
Warrier, Schiller, Frei, Haftel & Christner (2013)	Clinical Skills Examination Score	149	Clinical Skills Examination Score	668	*0.24*	*0.01*	Pediatric Clerkship	Stand. Exam	United States	Medical School
Warrier, Schiller, Frei, Haftel & Christner (2013)	Licensing Exam Score	311	Licensing Exam Score	614	*0.36*	*0.00*	Pediatric Clerkship	Stand. Exam	United States	Medical School
Warrier, Schiller, Frei, Haftel & Christner (2013)	Emergency Medicine Toddler Score	151	Emergency Medicine Toddler Score	281	*0.29*	*0.01*	Pediatric Clerkship	Course Exam	United States	Medical School
Warrier, Schiller, Frei, Haftel & Christner (2013)	Emergency Medicine Neonate Score	151	Emergency Medicine Neonate Score	281	*0.32*	*0.01*	Pediatric Clerkship	Course Exam	United States	Medical School
Wiener, Plass & Marz (2009)	Final Exam Score	220	Final Exam Score	1134	*0.57*	*0.01*	Functional Systems & Biological Regulation	Course Exam	Austria	Medical School
Willett, Rosevear, & Kim (2011)	Non-Endo/Rheum Score	83	Non-Endo/Rheum Score	83	*0.39*	*0.02*	Clinical Pathophysiology (Endocrinology & Rheumatology)	Course Exam	United States	Medical School
Willett, Rosevear, & Kim (2011)	Endo/Rheum Score	84	Endo/Rheum Score	83	*0.31*	*0.02*	Clinical Pathophysiology (Endocrinology & Rheumatology)	Course Exam	United States	Medical School
Zingone et al. (2010)	Course Grade	37	Course Grade	27	*1.10*	*0.07*	Ambulatory Care	Course Grade	United States	Pharmacy School

* Value estimated from other information in the studies, see text for details

APPENDIX B

R Syntax

```
#import data
tbl.complete.data <- read.table("Metaanalysis.Database.txt")


# variables for analysis
tbl.var <- c("Code", "Data.Code", "Res.Des.Code", "Co.Stat.Code", "Ana.Type.Code",
"Analysis.Code", "Comp.Group.Code", "Out.Meas.Code",  "N.1", "M.1", "SD.1", "N.2",
"M.2", "SD.2", "Course.Sub.Code", "Dom.v.For.Code", "Ed.Level.Code",
"TBL.Mod.Code", "Est.Code")

# collapse the Education level variable
tbl.complete.data$Ed.Level.Code <-
ifelse(tbl.complete.data$Ed.Level.Code>=4,4,tbl.complete.data$Ed.Level.Code)

# collapse the Course subject
tbl.complete.data$Course.Sub.Code <-
ifelse(tbl.complete.data$Course.Sub.Code>=4,4,tbl.complete.data$Course.Sub.Code)

# subset data to only include variables of interest
tbl.data <- tbl.complete.data[tbl.complete.data$Data.Code==1,tbl.var]

#load metafor package
library(metafor)

# calculate effect sizes and sampling variance
tbl.es <- escalc(measure="SMD", m1i=M.1, m2i=M.2, sd1i=SD.1, sd2i=SD.2, n1i=N.1,
n2i=N.2, data=tbl.data)

# add the studies that where we calculated ESs manually
tbl.es$yi[146]
tbl.es$yi[146] <- 1.15
tbl.es$vi[146]
tbl.es$vi[146] <- 0.13

# create study variable
tbl.data$study <- 1:nrow(tbl.data)
```

```
# baseline model---three level model
tbl.baseline.3l <- rma.mv(yi=yi, V=vi, data=tbl.es, method="REML", random = ~ 1 |
Code/study)
summary(tbl.baseline.3l)
AIC(tbl.baseline.3l)

# baseline model--random effects
tbl.baseline.rma <- rma(yi, vi, data=tbl.es, method="REML")
summary(tbl.baseline.rma)
AIC(tbl.baseline.rma)

# moderator models
# domestic vs foreign: Dom.v.For.Code
# outcome measure: Out.Meas.Code
# education level: Ed.Level.Code
# course subject: Course.Sub.Code

# dom vs. foreign
tbl.usa <- rma(yi=yi, vi=vi, data=tbl.es, method="REML", mods= ~
factor(Dom.v.For.Code))
summary(tbl.usa)
AIC(tbl.usa)

# outcome
tbl.outcome <- rma(yi=yi, vi=vi, data=tbl.es, method="REML", mods= ~
factor(Out.Meas.Code))
summary(tbl.outcome)
AIC(tbl.outcome)

# ed level
tbl.ed <- rma(yi=yi, vi=vi, data=tbl.es, method="REML", mods= ~
factor(Ed.Level.Code))
summary(tbl.ed)
AIC(tbl.ed)

# course subject: Course.Sub.Code
tbl.subj <- rma(yi=yi, vi=vi, data=tbl.es, method="REML", mods= ~
factor(Course.Sub.Code))
summary(tbl.subj)
AIC(tbl.subj)
```

```
# funnel plot
funnel(tbl.baseline.rma, yaxis="vi")

# funnel plot trimmed-and-filled
tbl.taf <- trimfill(tbl.baseline.rma, side="left")
funnel(tbl.taf)
```

BIBLIOGRAPHY

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716-723. doi:10.1109/TAC.1974.1100705

Anwar, K., Shaikh, A. A., Dash, N. R., & Khurshid, S. (2012). Comparing the efficacy of team based learning strategies in a problem based learning curriculum. *APMIS, 120*(9), 718-723. doi:10.1111/j.1600-0463.2012.02897.x

Beatty, S. J., Kelley, K. A., Metzger, A. H., Bellebaum, K. L., & Mcauley, J. W. (2009). Team-based Learning in Therapeutics Workshop Sessions. *American Journal of Pharmaceutical Education, 73*(6), 100. doi:10.5688/aj7306100

Bleske, B. E., Remington, T. L., Wells, T. D., Dorsch, M. P., Guthrie, S. K., Stumpf, J. L., . . . Tingen, J. M. (2014). Team-Based Learning to Improve Learning Outcomes in a Therapeutics Course Sequence. *American Journal of Pharmaceutical Education, 78*(1), 13. doi:10.5688/ajpe78113

Bloom, B. S. (1974). *Taxonomy of educational objectives; the classification of educational goals*. New York, NY: D. McKay.

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 221--235). New York, NY: Russell Sage Foundation.

Brandler, T. C., Laser, J., Williamson, A. K., Louie, J., & Esposito, M. J. (2014). Team-Based Learning in a Pathology Residency Training Program. *American Journal of Clinical Pathology, 142*(1), 23-28. doi:10.1309/ajcpb8t1dzkcmwut

Burgess, A. W., Mcgregor, D. M., & Mellis, C. M. (2014). Applying Established Guidelines to Team-Based Learning Programs in Medical Schools: A Systematic Review. *Academic Medicine, 89*(4), 678-688. doi:10.1097/acm.0000000000000162

Bushman, B. J., & Wells, G. L. (2001). Narrative Impressions of Literature: The Availability Bias and the Corrective Properties of Meta-Analytic Approaches. *Personality and Social Psychology Bulletin, 27*(9), 1123-1130. doi:10.1177/0146167201279005

Carmichael, J. (2009). Team-Based Learning Enhances Performance in Introductory Biology. *Journal of College Science Teaching, 38*(4), 54-61.

Cestone, C. M., Levine, R. E., & Lane, D. R. (2008). Peer assessment and evaluation in team-based learning. *New Directions for Teaching and Learning, 2008*(116), 69-78. doi:10.1002/tl.334

Chapman, J. G., Arenson, S., Carrigan, M. H., & Gryckiewicz, J. (1993). Motivational loss in small task groups: Free riding on a cognitive task. *Genetic, social, and general psychology monographs*.

Dickersin, K. (2005). Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In H. R. Rothstein, A. J. Sutton & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 11--33). Chichester, England: Wiley.

Dinan, F. J., & Frydrychowski, V. A. (1995). A Team Learning Method for Organic Chemistry. *Journal of Chemical Education, 72*(5), 429-431. doi:10.1021/ed072p429

Duval, S. J., & Tweedie, R. L. (2000a). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455–463.

Duval, S. J., & Tweedie, R. L. (2000b). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95, 89–98.

Egger, M., Jüni, P., Bartlett, C., Holenstein, F., & Sterne, J. A. C. (2003). How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technology Assessment, 7*, 1-68.

Farland, M. Z., Sicat, B. L., Franks, A. S., Pater, K. S., Medina, M. S., & Persky, A. M. (2013). Best Practices for Implementing Team-Based Learning in Pharmacy Education. *American Journal of Pharmaceutical Education, 77*(8), 177. doi:10.5688/ajpe778177

Fatmi, M., Hartling, L., Hillier, T., Campbell, S., & Oswald, A. E. (2013). The effectiveness of team-based learning on learning outcomes in health professions education: BEME Guide No. 30. *Medical Teacher, 35*(12). doi:10.3109/0142159x.2013.849802

Field, A. P. (2005). Is the meta-analysis of correlation coefficients accurate when population correlations vary? *Psychological Methods, 10*, 444-467. doi:10.1037/1082-989X.10.4.444

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology, 63*, 665-694. doi:10.1348/000711010X502733

Frash, R. E., Kline, S., & Stahura, J. M. (2004). Mitigating Social Loafing in Team-Based Learning. *Journal of Teaching in Travel & Tourism, 3*(4), 57-77. doi:10.1300/j172v03n04_04

Gelman, A. (2005). Analysis of variance: Why it is more important than ever. *The Annals of Statistics, 33*, 1-31.

Goldberg, H. R., & Dintzis, R. (2007). The positive impact of team-based virtual microscopy on student learning in physiology and histology. *Advances in Physiology Education, 31*(3), 261-265. doi:10.1152/advan.00125.2006

Haberyan, A. (2007). Team-Based Learning in an Industrial/Organizational Psychology Course. *North American Journal of Psychology, 9*(1), 143-152.

Haidet, P., Kubitz, K., & McCormack, W. T. (2014). Analysis of the team-based learning literature: TBL comes of age. *Journal on excellence in college teaching*, *25*(3-4), 303.

Haidet, P., Levine, R. E., Parmelee, D. X., Crow, S., Kennedy, F., Kelly, P. A., . . . Richards, B. F. (2012). Perspective: guidelines for reporting team-based learning activities in the medical and health sciences education literature. *Academic Medicine, 87*(3), 292-299. doi:10.1097/acm.0b013e318244759e

Hattie, J. (2012). *Visible learning for teachers* New York, NY: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81-112. doi:10.3102/003465430298487

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107-128. doi:10.3102/10769986006002107

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*, 486-504. doi:10.1037/1082-989X.3.4.486

Huggins, C. M., & Stamatel, J. P. (2015). An Exploratory Study Comparing the Effectiveness of Lecturing versus Team-based Learning. *Teaching Sociology, 43*(3), 227-235. doi:10.1177/0092055x15581929

Huitt, T. W., Killins, A., & Brooks, W. S. (2014). Team-based learning in the gross anatomy laboratory improves academic performance and students' attitudes toward teamwork. *Anatomical Sciences Education, 8*(2), 95-103. doi:10.1002/ase.1460

Ingram, V., Adams, J. S. (2003). Effects of Team Learning on Success Rates In Introductory Finance Classes. *Journal of Financial Education, 29*, 28-39.

Jarjoura, C., Tayeh, P. A., & Zgheib, N. K. (2014). Using Team-Based Learning to Teach Grade 7 Biology: Student satisfaction and improved performance. *Journal of Biological Education, 49*(4), 401-419. doi:10.1080/00219266.2014.967277

Johnson, J. F., Bell, E., Bottenberg, M., Eastman, D., Grady, S., Koenigsfeld, C., . . . Schirmer, L. (2014). A Multiyear Analysis of Team-Based Learning in a Pharmacotherapeutics Course. *American Journal of Pharmaceutical Education, 78*(7), 142. doi:10.5688/ajpe787142

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods, 17*, 137-152. doi:10.1037/a0028086

Kenny, P., McLaren, H., Blissenden, M., & Villios, S. (2015). Improving the Students' Tax Experience: A Team-Based Learning Approach for Undergraduate Accounting Students. *Journal of the Australasian Tax Teachers Association, 10*(1), 43-65.

Kent, S., Wanzek, J., Swanson, E. A., & Vaughn, S. (2015). Team-Based Learning for Students with High-Incidence Disabilities in High School Social Studies Classrooms. *Learning Disabilities Research & Practice, 30*(1), 3-14. doi:10.1111/ldrp.12048

Koles, P. G., Stolfi, A., Borges, N. J., Nelson, S., & Parmelee, D. X. (2010). The Impact of Team-Based Learning on Medical Students' Academic Performance. *Academic Medicine, 85*(11), 1739-1745. doi:10.1097/acm.0b013e3181f52bed

Koles, P., Nelson, S., Stolfi, A., Parmelee, D., & Destephen, D. (2005). Active learning in a Year 2 pathology curriculum. *Medical Education, 39*(10), 1045-1055. doi:10.1111/j.1365-2929.2005.02248.x

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods, 2*, 61-76. doi:10.1002/jrsm.35

Levine, R. E., O'boyle, M., Haidet, P., Lynn, D. J., Stone, M. M., Wolf, D. V., & Paniagua, F. A. (2004). Transforming a Clinical Clerkship with Team Learning. *Teaching and Learning in Medicine, 16*(3), 270-275. doi:10.1207/s15328015tlm1603_9

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

Malone, E., & Spieth, A. (2012). Team-based learning in a subsection of a veterinary course as compared to standard lectures. *Journal of the Scholarship of Teaching and Learning, 12*(3), 88-107.

McInerney, M. J., & Fink, L. D. (2003). Team-Based Learning Enhances Long-Term Retention and Critical Thinking in an Undergraduate Microbial Physiology Course. *Microbiology Education, 4*(1). doi:10.1128/jmbe.v4.68

Mennenga, H. A. (2012). Development and Psychometric Testing of the Team-Based Learning Student Assessment Instrument. *Nurse Educator, 37*(4), 168-172. doi:10.1097/nne.0b013e31825a87cc

Metoyer, S., Miller, S., Mount, J., & Westmoreland, S. (2014). Examples From the Trenches: Improving Student Learning in the Sciences Using Team-Based Learning. *Journal of College Science Teaching, 43*(05). doi:10.2505/4/jcst14_043_05_40

Michaelsen, L. K. (1992). Team learning: A comprehensive approach for harnessing the power of small groups in higher education. *To Improve the Academy,* 11(249).

Michaelsen, L. K., & Sweet, M. (2008). The essential elements of team-based learning. *New Directions for Teaching and Learning, 2008*(116), 7-27. doi:10.1002/tl.330

Michaelsen, L. K., & Sweet, M. (2011). Team-based learning. *New Directions for Teaching and Learning, 2011*(128), 41-51. doi:10.1002/tl.467

Michaelsen, L. K., Cragin, J. P., & Watson, W. E. (1981). Grading and Anxiety: A Strategy for Coping. *Journal of Management Education, 6*(1), 32-36. doi:10.1177/105256298100600107

Michaelsen, L. K., Knight, A. B., Fink, L. D. (Eds.) (2004). *Team based learning; A transformative use of small groups in college teaching.* Sterling, VA: Stylus.

Michaelsen, L. K., Watson, W. E., & Black, R. H. (1989). A realistic test of individual versus group consensus decision making. *Journal of Applied Psychology, 74*(5), 834-839. doi:10.1037/0021-9010.74.5.834

Mody, S. K., Kiley, J., Gawron, L., Garcia, P., & Hammond, C. (2013). Team-based learning: A novel approach to medical student education in family planning. *Contraception, 88*(2), 239-242. doi:10.1016/j.contraception.2012.07.012

Nicoll-Senft, J. (2009). Assessing the impact of team-based learning. *Journal on Excellence in College Teaching, 20*(2), 27-42.

Nieder, G. L., Parmelee, D. X., Stolfi, A., & Hudes, P. D. (2004). Team-based learning in a medical gross anatomy and embryology course. *Clinical Anatomy, 18*(1), 56-63. doi:10.1002/ca.20040

North, A. C., Linley, P. A., & Hargreaves, D. J. (2000). Social Loafing in a Co-operative Classroom Task. *Educational Psychology, 20*(4), 389-392. doi:10.1080/01443410020016635

Nyindo, M., Kitau, J., Lisasi, E., Kapanda, G., Matowo, J., Francis, P., & Bartlett, J. (2014). Introduction of team-based learning (TBL) at Kilimanjaro Christian Medical University College: Experience with the ectoparasites module. *Medical Teacher, 36*(4), 308-313. doi:10.3109/0142159x.2013.876490

Okubo, Y., Ishiguro, N., Suganuma, T., Nishikawa, T., Takubo, T., Kojimahara, N., . . . Yoshioka, T. (2012). Team-Based Learning, a Learning Strategy for Clinical Reasoning, in Students with Problem-Based Learning Tutorial Experiences. *The Tohoku Journal of Experimental Medicine, 227*(1), 23-29. doi:10.1620/tjem.227.23

Parmelee, D. X. (2010). Team-based learning: Moving forward in curriculum innovation: A commentary. *Medical Teacher, 32*(2), 105-107. doi:10.3109/01421590903548554

Parmelee, D. X., & Hudes, P. (2012). Team-based learning: A relevant strategy in health professionals' education. *Medical Teacher, 34*(5), 411-413. doi:10.3109/0142159x.2012.643267

Parmelee, D. X., & Michaelsen, L. K. (2010). Twelve tips for doing effective Team-Based Learning (TBL). *Medical Teacher, 32*(2), 118-122. doi:10.3109/01421590903548562

Parmelee, D., Michaelsen, L. K., Cook, S., & Hudes, P. D. (2012). Team-based learning: A practical guide: AMEE Guide No. 65. *Medical Teacher, 34*(5). doi:10.3109/0142159x.2012.651179

Persky, A. M. (2012). The Impact of Team-Based Learning on a Foundational Pharmacokinetics Course. *American Journal of Pharmaceutical Education, 76*(2), 31. doi:10.5688/ajpe76231

Persky, A. M., & Pollack, G. M. (2011). A Modified Team-Based Learning Physiology Course. *American Journal of Pharmaceutical Education, 75*(10), 204. doi:10.5688/ajpe7510204

Punja, D., Kalludi, S. N., Pai, K. M., Rao, R. K., & Dhar, M. (2014). Team-based learning as a teaching strategy for first-year medical students. *Australasian Medical Journal,* 490-499. doi:10.4066/amj.2014.2244

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Redwanski, J. (2012). Incorporating team-based learning in a drug information course covering tertiary literature. *Currents in Pharmacy Teaching and Learning, 4*(3), 202-206. doi:10.1016/j.cptl.2012.04.004

Robbins, S. P. (2000). *Managing Today!* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Roebuck, D. B. (1998). Using Team Learning in Business and Organizational Communication Classes. *Business Communication Quarterly, 61*(3), 35-49. doi:10.1177/108056999806100304

Searle, N. S., Haidet, P., Kelly, P. A., Schneider, V. F., Seidel, C. L., & Richards, B. F. (2003). Team Learning in Medical Education: Initial Experiences at Ten Institutions. *Academic Medicine, 78*(10). doi:10.1097/00001888-200310001-00018

Shankar, N., & Roopa, R. (2009). Evaluation of a modified team based learning method for teaching general embryology to 1st year medical graduate students. *Indian Journal of Medical Sciences, 63*(1), 4. doi:10.4103/0019-5359.49076

Sisk, R. J. (2011). Team-Based Learning: Systematic Research Review. *Journal of Nursing Education, 50*(12), 665-669. doi:10.3928/01484834-20111017-01

Stamatel, J. P., Bushway, S. D., & Roberson, W. D. (2013). Shaking Up Criminal Justice Education with Team-Based Learning. *Journal of Criminal Justice Education, 24*(3), 417-434. doi:10.1080/10511253.2013.782054

Sterne, J. A. C., Becker, B. J., & Egger, M. (2005). The funnel plot. In H. R. Rothstein, A. J. Sutton & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 75--98). Chichester, England: Wiley.

Sweet, M., & Michaelsen, L. K. (2007). How Group Dynamics Research Can Inform the Theory and Practice of Postsecondary Small Group Learning. *Educational Psychology Review, 19*(1), 31-47. doi:10.1007/s10648-006-9035-y

Tan, N. C., Kandiah, N., Chan, Y., Umapathi, T., Lee, S., & Tan, K. (2011). A controlled study of team-based learning for undergraduate clinical neurology education. *BMC Medical Education, 11*(1), 91. doi:10.1186/1472-6920-11-91

Thomas, P. A., & Bowen, C. W. (2011). A Controlled Trial of Team-Based Learning in an Ambulatory Medicine Clerkship for Medical Students. *Teaching and Learning in Medicine, 23*(1), 31-36. doi:10.1080/10401334.2011.536888

Thompson, B. M., Schneider, V. F., Haidet, P., Levine, R. E., Mcmahon, K. K., Perkowski, L. C., & Richards, B. F. (2007). Team-based learning at ten medical schools: Two years later. *Medical Education, 41*(3), 250-257. doi:10.1111/j.1365-2929.2006.02684.x

Tucker, J. M., & Brewster, M. P. (2015). Evaluating the Effectiveness of Team-Based Learning in Undergraduate Criminal Justice Courses. *Journal of Criminal Justice Education, 26*(4), 446-470. doi:10.1080/10511253.2015.1052000

Tuckman, B. W. (1965). Developmental sequence in small groups. *Psychological Bulletin, 63*(6), 384-399. doi:10.1037/h0022100

Valentine, J. C. (2012). Meta-analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 3: Data analysis and research publication* (pp. 485--499). Washington, DC: American Psychological Association.

Vasan, N. S., Defouw, D. O., & Holland, B. K. (2007). Modified use of team-based learning for effective delivery of medical gross anatomy and embryology. *Anatomical Sciences Education, 1*(1), 3-9. doi:10.1002/ase.5

Vaughn, S., Swanson, E. A., Roberts, G., Wanzek, J., Stillman-Spisak, S. J., Solis, M., & Simmons, D. (2013). Improving Reading Comprehension and Social Studies Knowledge in Middle School. *Reading Research Quarterly, 48*(1), 77-93. doi: 10.1002/rrq.039

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods, 10*, 428-443. doi:10.1037/1082-989X.10.4.428

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metaphor package. *Journal of Statistical Software, 36*(3), 1-48. URL http://www.jstatsoft.org/v36/i03/.

Wamsley, M. A., Julian, K. A., O'Sullivan, P., Mccance-Katz, E. F., Batki, S. L., Satre, D. D., & Satterfield, J. (2013). Team-Based Learning Exercise Efficiently Teaches Brief Intervention Skills to Medicine Residents. *Substance Abuse, 34*(4), 344-349. doi:10.1080/08897077.2013.787958

Wanzek, J., Kent, S. C., Vaughn, S., Swanson, E. A., Roberts, G., & Haynes, M. (2015). Implementing Team-Based Learning in Middle School Social Studies Classes. *The Journal of Educational Research, 108*(4), 331-344. doi:10.1080/00220671.2014.893224

Wanzek, J., Vaughn, S., Kent, S. C., Swanson, E. A., Roberts, G., Haynes, M., . . . Solis, M. (2014). The Effects of Team-Based Learning on Social Studies Knowledge Acquisition in High School. *Journal of Research on Educational Effectiveness, 7*(2), 183-204. doi:10.1080/19345747.2013.836765

Warrier, K. S., Schiller, J. H., Frei, N. R., Haftel, H. M., & Christner, J. G. (2013). Long-Term Gain After Team-Based Learning Experience in a Pediatric Clerkship. *Teaching and Learning in Medicine, 25*(4), 300-305. doi:10.1080/10401334.2013.827975

Wiener, H., Plass, H., & Marz, R. (2009). Team-based Learning in Intensive Course Format for First-year Medical Students. *Croatian Medical Journal, 50*(1), 69-76. doi:10.3325/cmj.2009.50.69

Willett, L. R., Rosevear, G. C., & 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? *Teaching and Learning in Medicine, 23*(1), 28-30. doi:10.1080/10401334.2011.536756

Wu, Y. P., Aylward, B. S., Roberts, M. C., & Evans, S. C. (2012). Searching the scientific literature: Implications for quantitative and qualitative reviews. *Clinical Psychology Review, 32*, 553-557. doi:10.1016/j.cpr.2012.06.007

Zingone, M. M., Franks, A. S., Guirguis, A. B., George, C. M., Howard-Thompson, A., & Heidel, R. E. (2010). Comparing Team-Based and Mixed Active-Learning Methods in an Ambulatory Care Elective Course. *American Journal of Pharmaceutical Education, 74*(9), 160. doi:10.5688/aj7409160