The development of team trust over time and its effect on performance when using Michaelsen's Team-Based Learning

by

Vanessa A. Preast

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Education (Curriculum and Instructional Technology)

Program of Study Committee:
Dale Niederhauser, Co-major Professor
Jared Danielson, Co-major Professor
Ana-Paula Correia
Holly Bender
Gary Phye

Iowa State University

Ames, Iowa

2012

Table of Contents

List of Tables

List of Figures

List of Nomenclature

TBL – Team-Based Learning
RAT – Readiness Assessment Test
iRAT –Readiness Assessment Test taken as an individual
gRAT –Readiness Assessment Test taken within the group
CDQ – Case Discussion Quiz
iCDQ – Case Discussion Quiz taken by individuals
gCDQ – Case Discussion Quiz taken by the group
TTQ – Thursday Team Learning Quiz
iTTQ – Thursday Team Learning Quiz taken by individuals
gTTQ – Thursday Team Learning Quiz taken by group
IF-AT –Immediate Feedback-Answer Technique

Abstract

Proponents of Michaelsen's Team-Based Learning (TBL) have claimed this teaching method quickly produces highly effective teams which are characterized by high trust among team members. Presumably, the high trust boosts performance because members feel less inhibited during discussions involving sharing personal views and challenging others' views. These team interactions can determine how well teams utilize their members' intellectual resources and make decisions. These assertions, while logical and compelling, did not appear to be fully verified within the TBL literature. This exploratory study used mixed qualitative and quantitative methods to describe performance, trust and behavior patterns over time within TBL teams in a second-year veterinary medical course. No variables were experimentally manipulated. Throughout the semester, I measured performance using students' individual and group quiz scores and measured trust using students' responses to custom trust surveys. Within the context of this study, there was no significant relationship between trust and performance and no consistent increase in trust over time. To investigate team behaviors, I observed five randomly-chosen teams on a weekly basis for the entire semester. The analysis suggested that some teams increased the speed and density of their intra-team communication and experienced greater member participation with time. The task type (in-lecture "clicker" questions vs. quiz questions) appeared to elicit different team interactions. The data hinted that decision-making may have been based on the majority answer held by members and that individual correct or incorrect members may have difficulty swaying the group. The students' open-ended responses illustrated fourteen characteristics that teams desired in their members. Although the findings are not generalizable, this study suggests a variety of interesting avenues for future study. For example, researchers may wish to explore how specific behaviors

exhibited by teammates influence group processes, decision-making, trust and performance within TBL.

Chapter 1. Introduction

Imagine walking on a college campus and asking random students, faculty and staff about their experiences working in groups. You would likely receive a wide range of responses from excitement to frustration. Most people recognize working in groups to be a necessity, especially in the workplace, and their personal experiences have shown some group experiences to be productive and rewarding and others to be a hindrance. At best, groups can produce outcomes far beyond the capabilities of any individual; but, at worst, group work can be deleterious to the group and individual members. Hackman (1983, p. 1456) described one situation exemplifying the best and worst aspects of groups: one workplace group "had established a production norm of fifty units a day" and so successfully discouraged one member's attempts to produce more such that this member's "output finally dropped even below the fifty-unit norm;" yet, when she no longer worked with that same group her "output soon *doubled*."

Researchers have documented conflicting evidence of group effectiveness. For example, Arum and Roksa (2011, p. 100) describe "a positive association between learning and time spent studying alone, but a negative association between learning and time spent studying with peers" despite other indications that "greater time spent working in groups leads to more favorable attitudes among students in general and that even minimal group work can have positive effects on student achievement" (Springer, Stanne, & Donovan, 1997, p. 23). Under the right conditions, group interactions may provide new ideas, insights, strategies, solutions or efforts not previously held by any individual member (Hill, 1982; Watson, Cooper, Torres, & Boyd, 2008; Watson, Johnson, & Merritt, 1998). Left alone, some cooperative learning groups will develop into more productive, positive learning environments for members while others will create more limited learning environments (Lisk, 2003; Wheelan & Lisk, 2000). The

inability to consistently provide equally beneficial group learning opportunities for all students may cause some instructors to abandon cooperative learning. By so doing, they also abandon the potential benefits of groups. Rather than avoiding groups in academic environments, instructors need guidance in establishing the conditions for forming and managing effective, well-developed learning groups. Not infrequently, "many well-intentioned faculty assign group projects without providing students the information and guidance prescribed by cooperative learning advocates" (Colbeck, Campbell, & Bjorklund, 2000, p. 61), and it is unsurprising when these haphazard approaches do not allow instructors or students to create the necessary conditions for optimal learning in groups within and outside of the classroom.

Hoping that a new cooperative learning method would help solve their teaching and learning problems, instructors from various fields have turned to Michaelsen's Team-Based Learning (TBL), which was described in the book *Team-based learning: A Transformative Use of Small Groups in College Teaching* (Michaelsen, Knight, & Fink, 2004). TBL, which promises to harness the power of well-functioning groups and avoid the problems of dysfunctional groups, is an increasingly popular instructional strategy among teachers in higher education and medical professional schools. Can these instructors feel justifiably confident that the system will meet their needs? Numerous case reports describing instructors' successful TBL implementations demonstrate that the technique is adaptable and can promote positive outcomes (Beatty, Kelley, Metzger, Bellebaum, & McAuley, 2009; Conway, Johnson, & Ripley, 2010; Dunaway, 2005; Feingold et al., 2008; Meeuwsen, 2003; Zgheib, Simaan, & Sabra, 2010). Empirical studies have pointed out that TBL teams can outperform their best member (Michaelsen, Watson, & Black, 1989; Watson, Michaelsen, & Sharp, 1991) and have examined how various factors, such as diversity or time, affect group processes and outcomes (Birmingham & Michaelsen, 1999; Watson, Johnson, Kumar, & Critelli, 1998; Watson, Johnson, &

Merritt, 1998; Watson, Kumar, & Michaelsen, 1993). In healthcare environments, TBL has produced equivalent or better learning outcomes than traditional educational interventions, including other small group learning methods (Naik, Teal, Rodriguez, & Haidet, 2011; Pileggi & O'Neill, 2008; Thomas & Bowen, 2011; Willett, Rosevear, & Kim, 2011). These studies tended to answer questions about whether or not the system worked, how well it worked, and how certain team characteristics affected outcomes. Unfortunately for instructors, the TBL literature explaining why this instructional strategy works and what happens among the members during team development is relatively sparse, providing little guidance for determining how and when to use this learning system.

The TBL proponents stress that a unique feature of TBL is its "reliance on the special characteristics of teams to accomplish a special kind of learning," and TBL can transform less effective "groups" into highly effective "teams." One of the two characteristics that distinguish "teams" from "groups" is that teams are characterized by "a high level of trust among the members of the group" (Michaelsen, et al., 2004, p. 12). TBL mechanisms supposedly promote trust development "through a series of positive group interactions," which then promotes "give-and-take discussion" (members freely sharing views and challenging each other's views) critical for effective group processes and learning. "By adhering to the Four Essential Principles of Team-Based Learning, teachers ensure that the vast majority of groups will develop a level of cohesiveness and trust required to transform groups into effective learning teams" (Michaelsen, et al., 2004, p. 35). As members see "other team members reliably complete tasks over time" (Michaelsen, et al., 2004, p. 77) trust develops and allows members to have the "kind of give-and-take discussion that is essential to group and team effectiveness, regardless of the setting" (Michaelsen, et al., 2004, p. 77). Trust develops because TBL creates the conditions in which members can have positive interactions, and this trust then encourages members to interact in

more effective ways which, in turn, promotes learning and builds greater trust to the point that these high-trust teams tend to perform at a high level.

Although TBL proponents identified trust development over time as an important underlying feature of effective teams, no TBL-specific studies appeared to confirm that the TBL mechanisms do promote trust-building with time and that this does alter the members' interaction patterns which in turn increase the team's effectiveness; nor did they measure trust, time, and team performance and their relationships to observed team interactions. Also, few, if any, TBL studies have utilized a formalized observation protocol for collecting data about teams' interaction patterns over time. Intrigued by this apparent gap in the literature, I sought to better understand how trust, performance and group interactions changed with time when using a Team-Based Learning format. In particular, I chose to investigate the presumption that, over time, TBL increases trust among team members, and that higher trust results in different group interaction patterns and better performance. My research questions included:

1. How does trust level relate to performance and time?
    a. Does trust predict group quiz performance?
    b. Does trust increase with time?

2. What is the relationship between team members' trust levels in their team mates and individuals' quiz performance?

3. What behavior patterns do teams exhibit in relation to TBL activities and time?

In the present study, I used mixed qualitative and quantitative methods to examine, on both team and individual levels, the relationship between trust level and learning outcomes within the context of a veterinary medical class using TBL. Learning outcomes were determined based on group and individual quiz scores (performance) in the class. Trust was defined as "the willingness of a party to be vulnerable to the actions of another party based on the expectation

that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (Mayer, Davis, & Schoorman, 1995, p. 712). Trust level was determined based on participants' responses on a trust survey. Both group performance and trust were expected to generally increase over time with the possible exception of temporary drops during periods of conflict. The results of this exploratory study may not be generalizable; however, they suggest future avenues for experimental investigations into why TBL and its sub-components produce particular outcomes so that instructors can have realistic expectations and make informed choices about utilizing TBL.

Chapter 2. Literature Review

A pragmatic approach to group learning shapes both my inquiry and interpretation of the literature because my ultimate goal is to understand which instructional changes lead to which learning outcomes. Teasing out how certain teaching techniques effect learning within an instructional system like TBL can be difficult due to its multifaceted nature. Not only does the system contain many different components and implementation details, but also each component likely interacts with many other factors including human social interactions, how people learn, time, course content, course design, and student characteristics. As a result, potentially relevant literature spans a wide range of topics, including trust, group development, human behavior, instructional design and human learning, each having potential for lifelong study. This review pulls educationally-relevant information from various fields in an attempt to draw practical conclusions for instructors. The topics include cooperative learning, time, group processes, and trust.

Cooperative learning

Various terms, including learning groups, collaborative learning, cooperative learning, and team-based learning, have been used to describe the same general idea of "putting individual students in a class into small groups for the purpose of promoting more active and more effective learning" (Fink, 2004, p. 5). Although Fink separates TBL from "cooperative learning," this study categorizes TBL as a highly structured subset of cooperative learning within a wide spectrum of cooperative learning styles. As such, the cooperative learning literature and small group teaching recommendations should generally apply to TBL. A number of studies show the learning benefits of cooperative learning across disciplines and cooperative learning strategies (Johnson, Johnson, & Smith, 1998; Skon, Johnson, & Johnson, 1981).

In practice, cooperative learning has several general characteristics: (a) a clear set of learning objectives that are accepted by all students, (b) clear and complete task-completion instructions, (c) heterogeneous groups which can interact face-to-face, (d) task structures ensuring positive interdependence, alignment between learning objectives and assessments, and relevant practice opportunities, (e) positive social interaction behavior and attitudes, (f) sufficient time spent on learning and task completion (at least four weeks in same group), (g) individual accountability and opportunity for success, (h) publically rewarded academic success, and (i) post-group reflection on effective group behaviors and attitudes (Johnson, et al., 1998; Lisk, 2003; Stahl, 1994).

Cooperative learning, however, does not guarantee success for all learning groups. Some groups develop successful learning behaviors, and others do not (Lisk, 2003; Watson & Michaelsen, 1988; Wheelan & Lisk, 2000). As a result, researchers have worked to devise ways to develop successful learning behaviors. In a number of settings, researchers have found that training helps group members learn and work together more effectively (Ashman & Gillies, 1997; Woolley, Gerbasi, Chabris, Kosslyn, & Hackman, 2008). TBL is designed to provide the benefits of well-functioning learning groups without the requirement of training. When the TBL activities are properly designed and implemented, students are expected to automatically develop good group processes which help them function effectively as teams (Fink, 2004, p. 13).

Michaelsen's Team-Based Learning (TBL). Team Based Learning is comprehensively described in the book *Team-Based Learning: A Transformative Use of Small Groups in College Teaching* (Michaelsen, et al., 2004). TBL is considered to be a particular instructional strategy with a specific combination and sequence of learning activities, not a series of independent small group activities (Fink, 2004). The strategy is described for a typical 15-week semester-long college course which is divided into five to seven units. Overall, four principles are

essential for successful TBL implementation: (a) instructors properly form and manage groups, (b) the process holds students accountable for individual and group work quality, (c) students receive frequent and timely feedback, and (d) assignment are designed to promote learning and team development (Michaelsen, 2004a; Michaelsen & Sweet, 2008). These principles are accomplished by incorporating the following nine characteristics into the course:

1. *Instructor forms diverse small groups.* The instructor assigns students to groups of five to seven people in such a way that relevant resources (students' characteristics, knowledge, background, culture or skills) are relatively evenly divided among the teams. The groups are still small enough for all members to participate.
2. *Instructor keeps teams together for the entire course.* Students in groups need time to get to know each other well enough to start functioning effectively as a team.
3. *Groups complete work during class.* All group members should be present for group activities, and it is difficult to coordinate all members outside of class. Also, groups may be more likely to divide up work among individuals when work occurs outside of class.
4. *Instructor does not assign roles within groups.* Teacher-assigned roles deprive students from the opportunity to learn how to manage group roles and functions effectively and efficiently.
5. *Groups complete graded individual quizzes followed by graded group quizzes.* Graded group work should constitute a significant percentage of the course grade because groups need an incentive for becoming an effective team. Members are motivated to contribute their best.

6. *Students receive immediate feedback.* Students are able to engage in analysis of what went wrong when they still can clearly remember the experience. The feedback helps them learn and improve their performance.
7. *Students provide graded peer assessments.* Individuals are discouraged from social loafing because peers can penalize freeloading individuals and reward contributing members.
8. *Groups simultaneously report specific decisions on the same problem.* Students are most engaged in discussion when they are wrestling with the same problem and are interested to know what other classmates decided.
9. *Instructor designs activities to encourage group decision-making.* The activities must not be easily sub-divided or completed by individuals in teams because that eliminates the group character of the assignment.

The TBL structure incorporates these nine elements within a three-phase sequence (preparation, application, and assessment) for each unit (Fink, 2004). The first phase introduces the students to the primary content material through out-of-class reading and in-class quizzes with feedback. The second phase increases the students' skills through homework and increasingly complex, in-class group assignments. The third phase assesses the students' knowledge through a culminating project or examination. These phases are detailed as follows:

When a new unit begins, students enter Phase I (preparation). They must complete out-of-class readings and come to class prepared to begin the "Readiness Assurance Process." In class, students take a short, often multiple-choice, test on the readings as individuals (iRAT). After students turn in their individual tests, the team members take the same test again as a group (gRAT), preferably using IF-AT (Immediate Feedback-Answer Technique) scratch-off cards, which provide immediate feedback. The individual and group tests are graded in class

and contribute to students' scores. The group scores are made public. Optionally, groups can try to regain deducted points by convincing the teacher to grant their written appeal in which they use reading assignments to support their answer choices. The preparation phase ends with the instructor offering "corrective instruction" during which the instructor, who now knows what ideas students were unable to understand on their own, offers any additional instruction deemed necessary for students to correctly understand the key concepts. The students, having already struggled with the material, are ready to listen closely to the instructor's brief, focused statements. "This process occurs five to seven times per course and constitutes the first set of in-class activities of each major instructional unit" (Michaelsen, 2004a, p. 41).

Phase II (application) begins with the assumption that the previous phase helped students acquire the declarative knowledge they needed to move forward. This phase develops students' procedural and strategic knowledge. Several class sessions are devoted to completing a series of meaningful small group application exercises in which teams use course content to formulate responses to increasingly difficult questions and problems. These activities must be designed so that the small groups all work on the same, significant problem or question and must generate a specific answer which all teams will share with the class simultaneously (Michaelsen, et al., 2004). Then, the instructor leads a discussion during which groups compare their responses, and the instructor offers feedback on the response quality. During this phase, a separate individual-portion-first component is not required, and teams may immediately begin work as a team. It is critical that the application phase utilizes "good application-focused group assignments [which] foster give-and-take discussions because they focus on decision making (not writing) and enable students to share their conclusions in a form that enables prompt cross-team comparisons and feedback" (Michaelsen, 2004a, p. 44).

Finally, the students move into Phase III (assessment) during which the students take an examination or complete one more group activity which contributes to their course grade. This phase determines how much the students learned during the instructional intervention. After completing this phase, the class moves into the next unit and the cycle repeats while integrating previous material with new course material.

Pure TBL incorporates all of the features described above. Yet, "teachers can and have borrowed 'pieces' of team-based learning (usually the 'individual-small group' sequence for testing)" (Fink, 2004, p. 9). Also, many studies describe various adaptations of TBL, including but not limited to eliminating phases, allowing students self-selecting groups, having more than seven students per team and not using Readiness Assurance, peer evaluations, or graded group work (Abdelkhalek, Hussein, Gibbs, & Hamdy, 2010; Conway, et al., 2010; Parmelee, 2010; Zgheib, et al., 2010). Michaelsen and Richards (2005) warned that educators who partially adopt the TBL system might underestimate the potential of team learning or might incorrectly attribute certain outcomes to TBL. TBL proponents believe that instructors may experience disappointing outcomes (e.g., negative experiences, student resistance, or some combination of feelings of indignation, frustration, and general distaste for the process) if they incompletely or inadequately enact TBL within their course (Lane, 2008). Several specific common misapplications of TBL include: a) overusing readiness assurance tests (measuring factual memorization) and underusing or poorly designing the assignments, b) omitting peer evaluations, and c) failing to properly introduce and acclimate students to TBL in the beginning (Michaelsen & Fink, 2008). Yet some researchers believe that the TBL method "allows flexibility for instructors to use selectively 1 or more of the phases, depending on the contextual demands of the course or session" and indicate that such flexibility is particularly important in medical education due to the unique constraints of the field (Thompson et al., 2007, p. 251). There has

been little research to indicate how these differences could impact the outcomes or what are the optimal factors associated with TBL use. For this reason, the course in this study was designed to adhere to the nine critical elements, and any deviations from pure TBL structure are described in detail.

Factors influencing group effectiveness. The literature has demonstrated how some of the components used by TBL affect group effectiveness. These include group diversity, size, and membership stability as well as task type and time.

Diverse groups. TBL uses an in-class team forming method to evenly distribute relevant member assets, liabilities, and characteristics (e.g., gender, ethnicity, work experience, relevant coursework, cultural perspectives, attitudes towards the course, intellectual ability, language fluency) equitably among the teams (Michaelsen, 2004a). Diversity is advantageous to groups because pooling member contributions (e.g., ability and experience) ensures that each group has the necessary resources for completing difficult tasks and working through individual differences helps teams utilize those member resources most effectively. Placing high- and low-ability (IQ) individuals together provides learning benefits for all individuals. Ten-year-old children working in mixed ability pairs performed better on post-tests than those working in homogeneous ability pairs or as individuals (Amaria, Biran, & Leith, 1969). When using TBL strategies in medical education, both the highest and lowest academic quartile students had improved performance, but the lower-quartile students had the greatest learning benefit (Koles, Stolfi, Borges, Nelson, & Parmelee, 2010). In both these situations, the efforts that the above-average individual expends when teaching the below-average individual results in helping each person learn. Furthermore, TBL placed “no burdensome duty” on higher-performing students because they “clarify their own knowledge by verbalizing and negotiating with peers, are rewarded with grades for their individual and team efforts, and spend no

additional time accomplishing those tasks beyond the live session" (Koles, et al., 2010, p. 1744). When working on long, complex tasks, the varied perspectives available in culturally diverse groups allows them to reach higher performance levels more quickly than non-diverse groups (Watson, Johnson, & Merritt, 1998). Although non-diverse groups working on less complex tasks initially performed better than diverse groups, over time the diverse groups overcame their process problems and improved their performance whereas non-diverse teams developed more individual difference problems (Watson, et al., 2008; Watson, Johnson, Kumar, et al., 1998; Watson, Johnson, & Merritt, 1998; Watson, et al., 1993). The obvious individual differences in diverse groups may have forced them to deal with and overcome those problems earlier in the process; however, the non-diverse teams may have initially ignored the less obvious individual differences, so they failed to develop a method of dealing with individual differences and individualistic activities began to negatively impact their effectiveness.

Group size. Group size can influence both group development and perceived group productivity. TBL recommends group sizes between five to seven members. This generally fits with the optimal group size range (three to six members) found in the literature (Birmingham & McCord, 2004; Bray, Kerr, & Atkin, 1978; Wheelan, 2009). When solving complex problems, groups need to be large enough to contain sufficient resources to be effective; yet, they must remain small enough to maintain a positive social atmosphere and avoid the coordination and developmental problems found in larger groups (Bray, et al., 1978; Hackman & Vidmar, 1970; Wheelan, 2009; Wheelan, Davidson, & Tilin, 2003).

Permanent team membership. In TBL, team member assignments are permanent. This feature may impart the stability and clearly identifiable membership which is important for team effectiveness. Team members and managers of long term work teams considered "boundedness" (clear boundaries distinguishing team members from non-members) and

stability of membership to be positively related to team performance (van Woerkom & Croon, 2009). Social worker teams which were clearer about who was or was not a member learned better, likely because the group was better able to coordinate decision-making and responsibility-sharing (Foldy & Buckley, 2010). Unclear membership can generate frustration and possibly undermine group effectiveness, especially when the questionable member was a supervisor who did not participate in group meetings but still "exercised a lot of authority over the team" (Foldy & Buckley, 2010, p. 44).

Assignment design. The TBL handbook, *Team-Based Learning: A Transformative Use of Small Groups in College Teaching*, devotes an entire chapter to creating effective assignments. Carefully attending to task design is very important because the characteristics of a task can influence the way the group approaches the task which, in turn, affects the outcomes (Hackman & Morris, 1975). Instructors can alter group processes and performance by changing the characteristics of a task, and analyzing the tasks to ensure they are appropriate for their intended outcome is an important step in troubleshooting problems during TBL implementation.

Due to the way individuals and groups approach a task, certain tasks are better suited for individuals and others for groups. Groups' ability to pool resources and correct errors help them make fewer indefensible hypotheses, categorize and prioritize information, and produce fewer repeated hypotheses than individuals (Hill, 1982; Laughlin, 1973). Thus, groups more successfully solve complex problems having multiple parts and no individual has all the necessary information (Hill, 1982; Watson, Johnson, & Merritt, 1998). However, on brainstorming tasks, pooling the ideas from individuals working separately produced more high-quality, unique ideas than groups, because the group interactions inhibited brainstorming performance (Hill, 1982). Easy tasks are not appropriate for group learning tasks because the

outcome is largely determined by one competent member and grouping people together merely increases the probability of having at least one competent member (Hill, 1982). Also, when tasks have readily recognizable solutions, higher ability (correct) members were less likely to be inhibited by lower-ability partners (Hill, 1982; Laughlin & Bitz, 1975). For more difficult or multi-step problems, the group was less reliant on particular individuals and benefited from the group's error-correction function and by having different members who could handle different steps of the problem-solving process (Hill, 1982). Since complex problems are subject to interpretation, the correct member may sometimes be overruled by the incorrect members; correct members were more likely to be overruled in low-ability groups (Hill, 1982).

Overall, individual efforts appear best for generating ideas and group efforts are best for integrating information and for clarifying and justifying evaluations. Task designs can optimize the individual and group benefits. For example, the Nominal Group Technique, which utilizes individual brainstorming followed by group discussion for clarification and evaluation followed by independent voting, was shown to generate significantly more unique ideas and greater participant satisfaction in the decision-making process than in interacting groups (Van De Ven & Delbecq, 1974). When utilizing iRATs followed by gRATs, TBL may be similarly capitalizing on both the individuals' superior idea generating capacity and the groups' superior error-correction ability. Additionally, TBL group activities should require members "to apply a rule or solve a problem" and to "make a concrete decision based on the analysis of a complex issue" (Michaelsen, Fink, & Knight, n.d., p. 6). Unlike "production tasks" intended to generate and present ideas, images or arrangements or "discussion tasks" involving evaluating issues and obtaining consensus, "problem solving tasks" should promote greater effectiveness, efficiency, teamwork and expression of opinions as members identify a course of action to resolve the problem (Hackman & Vidmar, 1970). This is probably because production tasks, such as writing

a story, are naturally individualistic activities, and their original, creative products with high literary quality come along with a negative tone, difficulty working together, and competitiveness possibly due to the need for divergent thinking (Hackman & Vidmar, 1970). TBL promotes convergent thinking, error correction and teamwork by avoiding tasks that are easily sub-divided, such as presentations and term papers. Instead, the team members must work together to make a specific choice, which they report simultaneously in a simple, succinct manner that can be directly compared to other groups (Michaelsen, et al., n.d.).

Hackman and Morris (1975) described an experiment which demonstrated how differences in the task design affected how groups approached that task. Groups were assigned to a process intervention (strategy instructions, anti-strategy instructions, and control) and a task condition (equal or unequal information distribution). Strategy groups were told to discuss their goals and how to work together, and anti-strategy groups were told not to waste time discussing procedures or strategy. Control groups received no special instructions. The teams were to assemble various electrical components based on a list of components' quantities and prices and were to maximize the total worth of the components they produced. Team members in the equal information groups received identical lists with item quantities and prices, but members of unequal information groups received lists with different quantities of the components that could be produced. The strategy groups performed well in the unequal task condition as did anti-strategy groups performing in the equal task condition. Interestingly, performance was lowered for the strategy groups in the equal task condition and the anti-strategy groups in the unequal task condition. Control groups were low for both. The strategy intervention helped improve group performance when the task actually required coordination and sharing among group members (i.e., unequal task); but, when the task could be done equally well without such coordination (i.e., equal task), strategy discussions deteriorated the

performance. Thus, for tasks that have an objective need to coordinate and share among group members (i.e., unequal task condition), it is beneficial for groups to engage in process planning discussions prior to completing the task. However, if the task can be completed well without coordinating and sharing among members (i.e., equal task condition), the groups might be better served by not wasting time with discussing strategies prior to engaging in the task. When designing assignments, instructors need to carefully consider how the task characteristics might influence the group processes, and they may need to implement instructional interventions to assist the groups with tasks that require coordination among members.

The task type ultimately influences the kind of learning that may occur. Various knowledge taxonomies have categorized the learning types to assist instructors with selecting appropriate tasks to accomplish the desired learning outcomes (Gagné, Briggs, & Wager, 1992; Phye, 1992; Phye, 1997a, 1997b, 1997c; Smith & Ragan, 2005). When applying Phye's functional knowledge taxonomy to TBL, it is evident that TBL components build upon each other to help students progressively attain declarative, procedural and strategic knowledge as they move towards increasing independent, problem solving ability. The activities in this study were designed to produce declarative, procedural and strategic knowledge (see Appendix B for class design details). The students demonstrated declarative knowledge when they completed a) the RATs, which tested recall of their reading material, b) the data abnormality part of Diagnostic Pathfinder homework cases, which required correctly spelled medical terms, and c) the within-lecture clicker questions. Students demonstrated procedural knowledge when generating diagnostic rationales for Diagnostic Pathfinder homework cases because the software guided them through the process of interpreting the cases. Students demonstrated strategic knowledge when they interpreted new cases without guidance and generated convincing rationales during quizzes and passed their final examination.

Practice tasks closely matching an authentic environment can help students successfully access prior knowledge when solving new problems in that environment. Students can better transfer their skills to a novel problem when the strategies used during training are compatible with the transfer task (Phye & Sanders, 1992). Students will learn what they practice, and they will practice what they are graded on (Phye, 1997b). For this reason, instructors who want students to organize their study strategies at a higher-order, problem-solving level must develop graded TBL activities that require students to practice critical thinking skills (procedural and strategic knowledge). The admonition to maintain consistency between the practice setting and "real world" also extends to the social environment. Students in an advanced 5th and 6th grade math class tested better individually if they had studied individually and they tested better as a group if they had studied as a group (Johnson, Johnson, & Scott, 1978). Fortunately, TBL has students practice their problem solving skills both individually and within groups all semester so they are prepared to successfully transfer their knowledge to interpreting novel cases they will encounter in veterinary practice.

Time. Classroom time is a precious commodity which must be used wisely to provide the greatest learning benefit possible. Both the time given to complete a particular task and the amount of time that groups remain together can influence the group effectiveness by affecting team processes, development, or productivity. Presumably, groups need more time than individuals to extract process gains from the time-consuming group discussion process. When given 50 minutes to write solutions to human relations problems, individuals had greater solution quality; but, individuals and ad hoc groups generated equivalent solutions when given 100 minutes (Fox & Lorge, 1962). TBL recommends using about 45 to 75 minutes for the Readiness Assurance process and about one to four hours for the application oriented activities

for each unit (Michaelsen, 2004a). In TBL, the large amount of time devoted to group activities probably allows TBL to extract the maximum process gains from the learning tasks.

Even more than time devoted to tasks, TBL proponents consider having the same members work together for the entire semester to be important for helping groups develop into high-functioning teams. "When the groups are properly formed, remain intact long enough to become cohesive teams, and are repeatedly given challenging tasks with prompt and clear feedback, then students learn the content, they learn how to use the content, they learn about themselves and how to interact with others on major tasks, and they learn how to keep on keep on learning after the course is over" (Fink, 2004, p. 8). Groups of various size appeared to have an inherent direction and sequence of development independent of its individual members (Wheelan & Krasick, 1993; Wheelan & Williams, 2003). Researchers have categorized these patterns into sequential, cyclic, life cycle, equilibrium, and adaptive/nonsequential group development models (Wheelan, 2005). Group interactions can change and become more effective over time; members of older work groups (together over five months) tended to perceive their groups to have more characteristics of the more productive development stages (Wheelan, et al., 2003). Culturally diverse groups, in particular, demonstrate changes in effectiveness over time such that newly formed groups have different performance characteristics from longer-term groups, and one should not expect newly formed, culturally diverse groups to solve problems very effectively (Watson, Johnson, Kumar, et al., 1998; Watson, Johnson, & Merritt, 1998; Watson, et al., 1993). Compared to long-term groups, newer group members: (a) have lower trust in and attraction to the group, (b) have little identification with the group or its goals, (c) are motivated by self-interest versus mutual support, (d) base their perceptions of others' skills on stereotypes versus actual behavioral observations, (e) use socially-focused decision-making behavior relying heavily on the best member instead of task-

focused communication, (f) have low willingness to disagree and focus on areas of agreement, (g) resolve conflict with less effective, face-saving strategies such as voting and compromise rather than consensus, and (h) are less able to complete difficult intellectual tasks due to low commitment to the group and avoiding open discussion (Birmingham & McCord, 2004; Sweet & Michaelsen, 2007).

With time, groups may pass through a conflict stage to reach a high-performance stage when the "group members comfortably and habitually share information with each other and have a relatively good sense of where different kinds of knowledge and expertise lie within the group" (Sweet & Michaelsen, 2007, p. 36). Work groups and academic groups in more advanced development stages were more productive, and more developed work groups tended to be those which were together longer (Lisk, 2003; Wheelan, et al., 2003; Wheelan & Lisk, 2000). Although we know that the amount of time together can affect group performance, we still do not know how many hours groups need to spend together to become productive. Some studies suggest that at least five to eight months may be necessary while others show groups performing effectively within hours. TBL teams require at least 20-25 hours working together "before they can fully assess and benefit from the resources of all members of the group" (Michaelsen, 2004a, p. 30; Watson, et al., 1991). Various factors ranging from the task type to group composition to environmental conditions may be influencing the amount of time needed to reach productive levels. It is possible that the structure of TBL may be promoting rapid group development, but researchers have not yet directly mapped TBL group development against existing models. Further research is necessary to clarify how groups progress through the group development phases over time when utilizing TBL.

Group process. The factors described above come together to influence the way that the members of a group work together to produce a group outcome. The group interaction

process can impact how member knowledge and skills are utilized for the task and can have lingering positive or negative effects on individual members' performance. According to Laughlin and Adamopoulos (1980), if individuals who were first incorrect when working individually later worked with a correct group, they had a probability of .80 of being correct in the final individual testing. However, initially correct individuals who subsequently were members in an incorrect group had a probability of .36 of being correct in the final testing. Working in a correct group is advantageous to previously incorrect individuals, but working in an incorrect group can be detrimental to a previously correct individual. When using group learning, educators want to take advantage of the first scenario but want to avoid the latter.

For tasks involving complex skills and high teamwork levels, the mediating role of group process is substantial thus increasing the potential process loss due to inadequately utilizing member skills and knowledge. Group processes also have potential to increase the total member talent available for completing the task. Rather than just gathering together what people already know, the group works to gain knowledge or generate skills that did not previously exist in the group. This increase in total talent (or process gain) is often an instructor's desired state when utilizing group learning and is presumably a benefit offered by TBL. The following may explain how TBL influences group processes to produce reliable effectiveness.

First, TBL may optimize group processes by enforcing social conditions most likely to result in productivity. Several cooperative learning theoretical frameworks emphasize the importance of designing learning tasks focused on a common goal (Meeuwsen & Pederson, 2006), and cooperative learning promotes a productive social environment by: (a) linking teammates' success to each other, (b) holding individuals accountable, (c) ensuring students promote teammates' success through encouragement, praise and assistance, (d) teaching

students social skills and how to use them properly, and (e) ensuring students take time to engage in group processing (Johnson, et al., 1998). Team-Based Learning formalizes each of these within its structure by utilizing gRATs, iRATS, group discussions with feedback and peer evaluations.

Second, TBL likely improves group effectiveness by adjusting three factors which can affect the interaction process. According to Hackman and Morris (1975), it is possible to improve group effectiveness by manipulating (a) group norms to improve performance strategies, (b) task design to influence member effort, and (c) group composition to influence how groups utilize and develop task-relevant knowledge and skill. TBL accomplishes each of these with its structure. Regarding group norms, the goal is to help members recognize and change existing norms which appear suboptimal for the task. TBL accomplishes this by having a structured system which rewards productive norms and inhibits disruptive norms. For example, TBL recommends giving practice RATs on the syllabus and then engage students in a structured grade weight setting exercise to discuss and decide how much the individual, group and peer evaluation scores would contribute to the grades. These lower-stakes activities allow members to begin trying out behaviors and discovering if those behaviors are acceptable to the group prior to working on real, graded assignments. Regarding task design, the goal is to ensure that tasks are exciting, fulfilling, and rewarding experiences so that members wish to exert effort in completing the task. TBL recommends that group activities require students to make a specific and simultaneous choice which can be compared among teams so that teams take ownership for their decision. Also, TBL group members receive individual scores, group scores and peer evaluation scores, so group members are held accountable for both their individual and group efforts which should reinforce “good” or “productive” behaviors and help individuals learn how well they are doing. Regarding group composition, the goal is to help teams

maximally utilize the talent in their teams. TBL addresses this by evenly distributing the member resources among the teams, keeping teams together all semester and providing immediate feedback on performance. This may provide a low-threat environment in which members can experiment with new forms of behavior and learn to trust each other.

Third, TBL utilizes methods that provide students with the most positive classroom experiences. Feichtner and Davis (1984) surveyed students at two universities to identify the most and least desirable cooperative learning qualities. The most positive experiences were characterized by instructor-formed groups, permanent groups, group sizes between four to seven members, and heterogeneous groups, plenty of in-class time for group work (averaged about 36% of class time), group work made up large portion of course grade (41-80%), peer grades are included as about 21 to 40% of the course grade, and students receive feedback on their performance. With a few minor differences, TBL recommendations are very similar. Interestingly, the students' perception of graded group work frequency depended on the task type. "Too few graded group assignments was detrimental to the group development process" did not give sufficient time for teams to become cohesive; yet, too many activities also had a negative affect which the authors presumed was due to student dividing work in an attempt to coordinate their efforts (Feichtner & Davis, 1984, p. 64). Students reported the best experiences in classes with one or fewer class presentations, two or fewer written reports and more than four group exams. These findings further validate TBL recommendations that (a) assignments which can be sub-divided, such as written reports, should be avoided, (b) group work should be performed in class to avoid coordination problems, and (c) reporting on team decisions should be specific and simultaneous, which is difficult with class presentations.

Finally, TBL is probably effective because the group assignment recommendations ensure good learning experiences for all students. In learning groups, members learn from each

other's expertise and acquire the information and skills they need as they work together to complete the tasks (produce the products) which provide the learning experience (Birmingham & McCord, 2004, p. 74; Colbeck, et al., 2000). This introduces an assessment problem because the group product quality is only a good measure of individual learning insofar as the group product demonstrates the individual effort and learning of team members. The assessments must accurately distinguish between team learning processes and their team performance outcomes because no change in performance might be incorrectly interpreted as no learning taking place (van Woerkom & Croon, 2009). The learning task also must be designed to provide practice opportunities on all aspects of the project for each individual. Subdividing labor to the most skilled group members can produce good products in both workplace and educational environments, but this evades the learning goals because students merely reinforce previous knowledge and learn little new (Colbeck, et al., 2000). Paradoxically, to promote the best learning experiences for members, learning groups might have to risk the project quality or efficiency as they involve their least skilled members in tasks outside their expertise. TBL avoids both the task design problem and the assessment problem. TBL requires that tasks not be sub-dividable so all members should be able to participate in the creation process. Also, TBL utilizes individual assessments (iRATs) in addition to group assessments.

Trust

Trust is one of the features distinguishing TBL "teams" from mere "groups" because teams have "a high level of trust among members of the group" (Fink, 2004, p. 12). Trust is the lubrication that allows team members to be "willing to challenge each other's views" and thus they can get the maximum learning benefit from engaging in give-and-take discussions (2004a, p. 35). In academic and lay literature discussions about group work, "trust" is often mentioned

as an important component for group effectiveness. Trust becomes a factor in group work because "working together often involves interdependence, and people must therefore depend on others in various ways to accomplish their personal and organizational goals" (Mayer, et al., 1995, p. 710). Arguments in the lay literature (Covey & Merrill, 2006) which claim trust affects the speed and cost of business have been echoed in the trust research. Tschannen-Moran and Hoy (2000) pointed out that distrust causes people to focus on self-protection, which increases costs in business and diverts energy from learning in academia. Organizational theory studies suggested higher trust is related to reduced negotiation costs and conflict levels and to superior information sharing, cooperation, and organizational performance (Gulati & Sytch, 2008). Even Wheelan's group development model (2005, p. 18) includes "trust and structure" as a stage in which groups have significantly increased capacity to work effectively and productively.

Since people claim trust is important, we would like to measure it. However, to measure trust, we must know what it is. Close examination reveals that "trust" is challenging. The word "trust" has been used as shorthand to describe a phenomenon that most people intuitively recognize as explaining good interpersonal results. Outside of the trust research field, researchers appear to default to a vague, ubiquitous, largely undefined, idea of trust that seems based on recognizing feelings from their own trusting experiences. Even when scrutinized by trust experts, trust is difficult to pinpoint, and trust researchers disagree about how to characterize and measure it. This introduces a problem when trying to determine if the "trust" described by the group process researchers is the same as the "trust" described by the trust researchers. As an analogy, a poet may speak of "sunlight," having experienced the brightness, beauty and warmth of the sun, but a physicist defines sunlight in mathematical formulas and describes it as waves and particles.

Perspectives on trust. The TBL literature and non-trust researchers (lay persons, group process researchers and group development researchers) appear to mostly use the shorthand, intuitive meanings of trust within their frameworks. These frameworks tend to focus on the emotional undertones that exist within the interpersonal interactions. They associate confidence and security with trust and suspicion with distrust (Covey & Merrill, 2006; Wheelan, 2005). A trust definition relying heavily on human emotion and attitudes can be difficult to measure and may not look the same from context to context. For example, "Trust issues are generally dealt with in more covert ways (such as joking, teasing, or testing) in task groups and in more overt ways in learning and therapy groups" (Wheelan, 2005, p. 14). The TBL literature tends to associate increased trust with increased willingness to engage in discussions and express differing opinions without expressly defining trust. Ultimately, using the shorthand "trust" might contribute to confusing comparisons across studies if different researchers utilize different implicit meanings and expressions of trust, but fail to explicitly define those meanings. Perhaps we think we're all talking about the same thing and really are not.

Within the realm of trust research, the literature is highly diverse and each researcher's approach partially describes trust and mechanisms leading to enhanced trust. Trust is a complex situational and multidimensional construct which is "activated and sustained by a multidimensional set of conditions;" and not only do researchers disagree about what these trust conditions are, but also "there is no instrument for measuring an exhaustive set of them" (Butler, 1991, pp. 644, 647). Viewed through the perspective of their various disciplines, trust researchers have alternatively viewed trust (a) as an individual characteristic (personality psychologists), (b) as expectations about others' behavior during transactions while focusing on the contextual factors influencing trust development and maintenance (social psychologists), or

(c) in light of the institutions and incentives created to increase trust by reducing the anxiety and uncertainty associated with transactions among relative strangers (economists and sociologists) (Bhattacharya, Devinney, & Pillutla, 1998). Economists also equate trust with people behaving in a prescribed manner, even if the behavior is related to how contracts or rewards and punishments are structured, because economists are more concerned about the costs and benefits of certain behaviors than about inherent trustworthiness. These differences can result in opposing views about the definition of trust. For example, trust researchers disagree about whether or not trust can be present when external mitigating factors, such as incentives or sanctions, are used to enforce desired behavior. This is a particularly important distinction for Team-Based Learning, since the system's structure is supposed to form groups into high-functioning, trusting learning teams.

Some trust researchers feel that the diversity of viewpoints may not be able to be converged. Having reviewed empirical generalized trust research, Nannestad (2008, p. 432) called it a "huge puzzle that is not even near solution" and concluded that we have little useful knowledge about generalized trust and how it explains social behavior, institutions or social and political change, but Nannestad admitted that recent trust research has identified pieces that do and do not belong to that puzzle. It is possible that "a universal conceptualization of trust and distrust may have difficulty attaining a sufficient level of theoretical and empirical viability for research purposes," and that "attempts to 'stretch' trust to cover the extant range of usages seem to be at an extremely high risk of producing constructions that are either too elaborate for theoretical purposes or relatively meaningless in the realm of empirical observation" (Bigley & Pearce, 1998, p. 408). Even if researchers were to agree upon a common definition of trust, "the reality of trust will not be the same at all times and in all places" (Tschannen-Moran & Hoy, 2000, p. 558) due to the variables involved. The characteristics of the

trustor and trustee as well as the context of the trusting relationship will determine how trust will play out in any situation.

Trust researchers have offered various schemes to make sense of the wide range of trust research. Most tend to view trust in relation to (a) individual trustor (the entity doing the trusting) characteristics, (b) trustee (the entity being trusted) characteristics, or (c) characteristics of the environment, relationship, or context. Within TBL teams, each member has influence on the group output and on other members; thus, each member acts as both trustee and trustor. Additionally, the TBL framework provides a context which modulates the trusting relationships. Since there is no consensus approach to trust, this study considers trustor, trustee, and contextual factors to all be important for a trusting situation. The framework used in this study resembles this formula: Trusting situation = trustee characteristics x trustor characteristics x contextual factors. Each of these categories is discussed below.

Trustor characteristics. Theorists focusing on trustor characteristics recognize that "people differ in their inherent propensity to trust" another before they have any information about the other party (Mayer, et al., 1995, p. 715). This propensity is this person's "general willingness to trust others" (Mayer, et al., 1995, p. 715). An individuals' values, attitudes, moods and emotions can affect their level of trust in a situation (Jones & George, 1998). Some research describes an individual's propensity for trust as a stable trait across situations, but others see it as situation specific with effects from personality and situational factors.

Trustee characteristics. Theorists who view trust "as an expectation that is specific to a transaction and the person with whom one is transacting" tend to see "vulnerability is a key element of trust" (Bhattacharya, et al., 1998, p. 460). Vulnerability implies something important can be lost; thus, vulnerability is risky. For this reason, the researchers focus on the perceived

trustee characteristics which make this entity worthy of the risk. Some of the trustee characteristics that might lead the trustor to trust the trustee include: availability, competence, consistency, discreetness, fairness, integrity, loyalty, honesty, openness, promise fulfillment, receptivity, benevolence, and reliability (Butler, 1991; Tschannen-Moran & Hoy, 2000). Mayer et al. (1995) summarized trustworthiness into three antecedents of trust, namely ability, benevolence and integrity, and noted that these factors can vary along a continuum.

Context characteristics. Theorists focusing on trust as a relationship may envision trust as emerging from prior contact and developing "knowledge-based trust" based on ongoing interaction and on learning more about the partner (Gulati, 1995). The trust experience can be determined by the interplay between simultaneously operating values, attitudes, moods and emotions "to produce an overall state of trust or distrust," so trust can be a dynamic experience shifting between distrust, conditional trust and unconditional trust (Jones & George, 1998, p. 537).The characteristics of the relationship partners can influence the relationship. For example, more similar partners "can derive greater stocks of trust from joint history compared to more heterogeneous sets of partners" (Gulati & Sytch, 2008, p. 182). It is important to determine how the parties establish the normative guidelines by which they act within the relationship. The norms can come from within or be influenced by external sources, but researchers disagree about how this relates to trust. Gulati (1995, p. 93) implies that an externally imposed governance structure is distinct from trust: "A detailed contract is one mechanism for making behavior predictable, and another is trust"; so when trust exists "people may not choose to rely upon detailed contracts to ensure predictability". Yet, this distinction may be less an either-or condition and more like a continuum because "cautious contracting gives way to looser practices as partner firms build confidence in each other…[so] familiarity between organizations through prior alliances does indeed breed trust" (Gulati, 1995, p. 105).

Sheppard and Sherman (1998) proposed another view in which trust takes on four different forms depending on the relationship form and depth. The relationships may be shallow or deep and have dependent or interdependent forms. Each relationship category carries different risks for the trustor, and the risks are additive from one category to the next. For example "deep interdependence" has all the risks of "shallow dependence," "deep dependence," and "shallow interdependence". In this model, trust is the "acceptance of the risks associated with the type and depth of the interdependence inherent in a given relationship," so trust exists at several levels with a clear order: "shallow dependence < deep dependence, or shallow interdependence < deep interdependence" (Sheppard & Sherman, 1998, p. 426). Because trust is a relational act that "involves the belief that features of the other, the relationship, or the context in which the relationship is embedded will mitigate the risks associated with that relational form," it is possible to enhance trust by selecting a trustworthy partner and utilizing institutional mechanisms to create a basis of trust between parties unable to determine trustworthiness themselves (Sheppard & Sherman, 1998, p. 426).

Trust in TBL. Ultimately, the trust literature tends to include the following themes about trust: (a) "trust exists in an uncertain and risky environment", (b) "trust reflects an aspect of predictability" and is an expectancy which includes some possibility of error , (c) the degree of trust is directly related to the magnitude of the expectancy, (d) the strength of trust is related to an entity's confidence in the trust, (e) "trust exists in an environment of mutuality" and is "situation and person specific", and (f) trust is related to good outcomes (Bhattacharya, et al., 1998, pp. 461, 462). The uncertainty and risk are familiar to students in cooperative learning groups who come to a new class, important for their academic programs, and find that their academic success is dependent on working with people they had never met. The TBL structure should support shallow dependent and shallow interdependent relationships within the

framework of Sheppard and Sherman's (1998) trust model because its built-in, trust-promoting features mitigate the risks associated with the cooperative learning relationship and thereby enhance trust. Shallow dependent relationships exist when TBL encourages reliability and discretion through deterrence (peer evaluations) and the inability to easily leave the relationship (course-long teams). The peer evaluations can penalize parties who perform unreliably or act in disagreeable ways. Shallow interdependent relationships exist when TBL reinforces consistency, transparency, and predictability by forcing team members to engage in discussion and attempt to understand each other's rationales for group work. Although TBL structures do not necessarily support deep relationships, it is possible some teams reach a level of trust which involves a sense of caring and benevolence towards teammates and having a sense of obligation towards the group. For example, a deep dependent relationship is more evident when a teammate trusts the team to make good choices on group activities during times when he or she must be absent. It is also evident when teammates feel obligated to come to class and participate for their team's sake, not just their own grades, or when they try to help each other understand and learn the material. It is unlikely there is deep interdependent relationship between teammates on TBL teams.

From an individual perspective within a team, especially at the beginning, members may perceive their outcomes (grades) to be dependent on the other members, so each student (trustor) likely would consider the other students' (trustee) characteristics (such as consistency and ability) when determining whose arguments to believe during group activities. Thus, TBL teams likely begin as "shallow dependence" relationships because: a) team members must rely on team mates to complete the group work, but have no direct control over another's behavior, and b) TBL mechanisms ensure the reliability, competence and integrity of the team members. In a dependent relationship, the trustor's outcomes are contingent upon the trustee's actions, so

it is a unidirectional dependency. Mayer (1995, p. 712) defines trust as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party". This definition describes well the "dependent" forms of trust likely to be found in academic groups, and thus formed the trust definition used in this study. If the trust level surpasses the perceived risk threshold, the trustor will engage in risk within a relationship but "the form of the risk taking depends on the situation" (Mayer, et al., 1995, p. 724). When determining trust level, it is not enough to ask "do you trust them?" as this must be further qualified with the question "trust them to do what?" (Mayer, et al., 1995, p. 729). Thus, this study utilized a trust measurement instrument that asked members to indicate how willing they were to rely on each other for contributing positively to the current group activity and allowed them to comment on their choice. The intention was to capture both the team's collective trust level, but also to indicate what individuals' trust levels were towards a particular trustee within a certain context.

In his opinion paper, Maccoby (2003, p. 60) explained that an institution can "avoid distrust by enforcing the rules, but it will build trust only by practicing transparency and increasing participation", and this belief echoes Michaelsen's ideas and the TBL structure. From this perspective, creating a trust culture involves creating a learning environment in which people are not afraid to make mistakes and they accept the inevitability of constant change. Maccoby suggested principles for building a trusting culture, and these can all be found within the TBL structure. TBL utilizes a distinct and clearly articulated structure for the class which includes the ability of students to write appeals and to participate in establishing grade weights. Thus, everyone (a) knows the goals and strategy, (b) knows the rules and can question

inconsistent decisions, (c) "knows how both the team and individual performance are scored and rewarded," and (d) participants "feel they are being treated fairly" (Maccoby, 2003, p. 60).

TBL establishes the conditions in which teammates can demonstrate, over time, their reliable contribution to the group, thus helping trust grow. TBL should reinforce individual accountability and learning as well as establish effective working relationships with teammates. Underprepared students would face reproach from other teammates which creates pressure to be prepared for graded assignments in the future. Non-contributing students would risk lower scores due to the peer evaluations and are thus motivated to work harder. Immediate feedback would show team members when they failed to best utilize resources within the team, so members become motivated to figure out how to best utilize the strengths of individual members. In TBL, team members take social and academic risks when they share their opinions during group discussions; presumably, members who trust others in the team would be more likely to take the risk to speak up or to listen to others in the team.

Influence of time. "A number of theorists have suggested that trust evolves over time based on a series of observations and interactions" (Mayer, et al., 1995, p. 730). Theoretically "trust can be utilized and hence demonstrated and reinforced only in situations of risk and uncertainty" thus "long-lasting histories of interaction provide unparalleled opportunities for building mutually trusting relationships" (Gulati & Sytch, 2008, p. 168). Although Gulati & Sytch (2008, pp. 179, 181) found an "absence of support for the direct effects of history on trust," they also found a "non-linear relationship between past history and interorganizational trust". They theorized that the early stages in the relationship provide limited opportunity to develop trust and that relationships go through a period of ambivalence in which the parties are assessing the trustworthiness of the partner. Only later, after a certain threshold in the relationship, do the partners reap the benefits of trust. The trust level evolves and fluctuates based on interactions

between the parties: “When a trustor takes a risk in a trustee that leads to a positive outcome, the trustor's perceptions of the trustee are enhanced” but unfavorable conclusions would cause the trustor’s perception of the trustee to decline (Mayer, et al., 1995, p. 728).

A different perspective about the evolution of trust involves a temporary assumption of trust during a newly-formed relationship. When beginning a new relationship and as long as there is no obvious sense that the parties’ values are so divergent that one might be vulnerable, each party temporarily “suspends belief that the other is not trustworthy and behaves as if the other has similar values and can be trusted” (Jones & George, 1998, p. 535). Beginning with trust is advantageous because it is easier (consumes less time and resources) than beginning with distrust. The outcomes of the initial encounters then influence the future trust. “Successful behavioral exchanges are accompanied by positive moods and emotions, which help to cement the experience of trust and set the scene for the continuing exchange and building of greater trust” (Jones & George, 1998, p. 536). Their model depends on the parties jointly creating the social situation, trying to understand the other’s expectations, needs and goals and adjusting communication and behavior “to fit the unfolding, mutually determined definition of the social situation” (Jones & George, 1998, p. 535). Eventually, the willing suspension of disbelief relating to the other’s trustworthiness transforms into the desire to trust each other “because both feel secure that they will not be harmed or put at risk by the actions of the other and that the other is, in fact, trustworthy” (Jones & George, 1998, p. 536). To experience an evolution in trust, the partners must have (a) “strong confidence in each other’s values and trustworthiness”, (b) “favorable attitudes towards each other” and (c) have positive feelings about the relationship (Jones & George, 1998, p. 536).

The presumption that TBL helps teams increase trust over time is supported by the trust theorists. The TBL structure provides “a series of interactions that enable individual

members to test the extent to which they can trust their peers to take them seriously and treat them fairly" (Michaelsen, et al., 2004, p. 30). Therefore, as TBL helps individuals to demonstrate their trustworthiness in the form of reliable contributions to the group, the trust in the groups builds over time.

Influence on performance. Teams in later developmental stages (stages three and four) performed better, and member trust and cooperation was an integral part of Wheelan's (2005) third stage, called "Trust and Structure." Teams in this stage are beginning to be productive. Before they can reach this phase, though, teams first pass through a conflict phase. Constructive conflict can be helpful for performance, yet is potentially detrimental to the social bonds within a group, so newly-formed teams might opt for a less-effective, but less threatening, compromise strategy (Birmingham & Michaelsen, 1999) . With time, teams utilized compromise less frequently, possibly because the group trust had increased over time, the groups had more practice with constructively handling conflict, and they needed time to build sufficient trust for members to be comfortable with open disagreement (Birmingham & Michaelsen, 1999).

Examining virtual and face-to-face teams, both had initial high trust, but a more enduring trust had to be "maintained by positive, task-oriented team dynamics" involving teammates meeting work expectations. Face-to-face and virtual teams' "average enduring trust is highly positively related to the team performance, confirming the general consensus that higher trust teams perform better" (Corbitt, Gardiner, & Wright, 2004, p. 6). The authors concluded that "trust is related to what a person does on the team" and that team members "need to meet work expectations early in order to maintain the trusting environment" (Corbitt, et al., 2004, p. 7).

Jones and George (1998) analyzed how trust evolved in organizations and how it influenced cooperation and teamwork, which are important for organizational performance.

The authors categorized trusting states into conditional trust, unconditional trust and distrust. The authors proposed that unconditional trust, in which the parties' shared values create a common bond, can "fundamentally change the quality of the exchange relationship and convert a group into a team" (Jones & George, 1998, p. 539). The shared values and positive emotions accompanying unconditional trust manifest themselves as strong desires for team members to contribute to the common good. Unconditional trust promotes seven social processes leading to synergistic team relationships which lead to superior performance: "1) broad role definitions, 2) communal relationships, 3) high confidence in others, 4) help-seeking behavior, 5) free exchange of knowledge and information, 6) subjugation of personal needs and ego for the greater common good, and 7) high involvement" (Jones & George, 1998, p. 540).

The distinction Jones and George (1998) make between a group and a team echoes the difference proposed in *Team-based learning: A Transformative Use of Small Groups* between teams in TBL and groups in other, potentially less effective, collaborative learning methods. From this perspective, one could then argue that TBL is able to create highly effective and trusting learning teams because the structure of TBL promotes trust and trust makes teams more effective than groups. Based on trust theorists describing increasing trust over time and group developmental researchers describing increasing performance with stages, we might expect trust in TBL teams to start out moderately high (temporary assumed trust) then drop during the conflict phase then steadily increase as the interactions become more consistently positive. The trust studies do suggest a connection between greater trust and greater performance as asserted by TBL proponents. Thus, we should also expect to see the group performance pattern to resemble that of the group trust level.

Conclusion

There is still much to learn about what makes groups productive and how to harness the power of groups. Groups are unpredictable, complex, adaptive systems in which local interactions may generate global patterns which in turn influence local interactions. As members interact a global pattern may emerge and strengthen which then influences local member interactions to conform to the emergent global pattern (Wheelan & Williams, 2003). Group processes should be examined on three levels including, group interaction (individual dynamics), group development (larger-scale patterns), and group response to changing environmental constraints (Arrow, Poole, Henry, Wheelan, & Moreland, 2004) which, in TBL, would correspond with individual student member behaviors, the team's collective behavior, and the way TBL is implemented in the classroom. Within a particular TBL classroom, one would expect a limited number of group patterns to emerge since all students are experiencing the same larger contextual environment and the differences between teams lies primarily within the individual member differences. The same may not be as likely when comparing TBL implementations in different classrooms because the contextual variables have a non-linear impact on global behavior such that "large changes at some ranges of values will have no discernable [sic] effect, whereas small changes at some other ranges of values (near critical thresholds) will have dramatic effects on the group" (Arrow, et al., 2004, p. 97). Ultimately, "identifying the thresholds at which abrupt changes are likely should help us design more effective interventions" (Arrow, et al., 2004, p. 97).

Trust, time and performance are related in complex ways which have not entirely been worked out in the literature. Trust and time can influence performance, and performance and time can influence trust. The group trust and performance relationship is like a dance. The lead dancer does not forcefully move the partner in the desired direction; rather, leading is an offer

and the partner can choose to take the offer. When dancing, both persons remain attentive to the others' subtle body movements and move in relation to those cues. Such is the emotional and psychological give and take during trust building as outlined by Jones and George (1998). Both parties enter the dance with a certain amount of openness and receptivity to how the partner responds. With time, the partners "test" the relationship, and each positive outcome suggests that the dance partners can attempt increasingly challenging moves. Team Based Learning becomes the dance instructor whose guidance helps the partners work together better and simultaneously improve their technical skills.

Chapter 3. Methodology

This exploratory study was performed during an existing 15-week veterinary course using an adaptation of the TBL instructional methods described in the book *Team-Based Learning: A Transformative Use of Small Groups in College Teaching* (Michaelsen, et al., 2004). Because the TBL in this course was modified to fit the particular instructional context and learner preferences, the teaching methods used in this course are explained in detail in Appendix B to ease comparison of this TBL implementation with that of past and future studies.

Course Design

General course description. The course was a comprehensive introduction to veterinary clinical pathology which was hosted by an AVMA accredited veterinary school at a large Midwestern university. The course was attended by 145 second-year veterinary students from two universities: 121 local students attended in person and 24 distant students participated through teleconferencing technologies. The local and distant students could see and hear each other and the instructor.

The class, as a whole, consisted of 111 (76%) female and 34 (23%) male students, which fit into the typical 70-80% female typical of veterinary classes during this time. The students were divided into 24 permanent teams with six or seven members. Local and distant student teams participated in class activities by using either a radio frequency signaling device (a "clicker") or ResponseWare software to select their team answers. All team selections were collected by a central computerized system. Additionally, each student had a portable personal tablet computer which they used in class for online quizzes and note-taking.

The instructor maintained a WebCT site through which the students could access copies of lecture PowerPoints, lecture captures (video) of previous classes, class schedules, online quizzes and other pertinent information. The course was taught for a 50 minute session

Monday, Tuesday, Wednesday, and Friday and for a 100-minute session with a 10 minute break on Thursday. The class had approximately nine units which lasted from two days to two weeks depending on the topic. The instructor incorporated various interactive teaching techniques into most class periods including lectures peppered with clicker questions, food metaphors and skits, Case Discussion Quizzes (CDQs), and Thursday Team Learning Quizzes (TTQs). Each teaching technique is detailed in Appendix B.

The students' course grades came from points obtained by completing homework cases using the Diagnostic Pathfinder software, individual quizzes, team quizzes, clicker questions, final peer evaluation scores, and the final examination. The instructor predetermined that 50% of the students' grades would come from the final exam (1/3 of grade) and the Diagnostic Pathfinder homework (1/6 of grade). Students used consensus decision-making during a "Grade-Weight-Setting Exercise" (Michaelsen; Michaelsen, et al., 2004, pp. 241-248) to weight the remaining 50% of their grade from among their individual work, teamwork, and peer evaluations as 25%, 15%, and 10% respectively.

Compared to prior or subsequent courses this instructor had taught, the course in this study had some unusual features which potentially could have influenced the students' experiences. First, this was the first year the course included distant teams interacting with the classroom in real time through videoconferencing technology. Second, the instructor was in negotiations with a book publisher to make available an electronic textbook, but the deal was not finalized until midway through the semester.

Classroom environment. The classroom had tiered, stadium-like seating with chairs that could turn. Teams' assigned locations were arranged so that half of each team sat in a row in front of the other half. This allowed the front team members to turn and face their team members in the row behind them when they did group work.

The instructor had taught Clinical Pathology to veterinary students for 30 years. Data collection for this study occurred the 6th time she taught her cause using Team-Based Learning, so the instructor was using a stable and well-established adaptation of the Team Based Learning method.

Fidelity of TBL implementation. Because the course design adhered to the nine core characteristics of TBL and avoided the three major pitfalls identified by Michaelsen and Fink (2008), I considered this to be an authentic implementation of TBL. The instructor did deviate somewhat from a pure TBL process in the following two minor ways:

First, this course did not have three distinct phases within each unit. Instead, this course utilized a series of increasingly complex individual and group quizzes delivered using the Readiness Assurance Process format. These quizzes served different functions at different times. The earlier quizzes ensured that students had the necessary declarative knowledge to proceed, consistent with Phase I. Later, more challenging quizzes provided group practice opportunities, as expected during Phase II. Because the quizzes contributed to 40% of student grades, they provided the assessment of student knowledge typically found in Phase III.

Second, the instructor continued to use in-class lectures to introduce content and supplement out-of-class readings and homework even though TBL proponents recommend offloading lecture material into out-of-class readings to save in-class time for group activities. However, because students were in class every day of the week, with a double session on Thursdays, students had at least twice the amount of class time in this course than in a typical two-or-three-days-per-week course, which left plenty of time for both lecture and group work.

Participant Selection

I analyzed student-generated data from 136 students (90% of the class) who signed the informed consent document indicating willingness to participate in the study. This means that

individual quiz scores, survey responses and peer evaluation responses were removed for any non-consenting individual prior to analysis.

Teams eligible to be included in the observation selection pool met the following criteria: 1) the team members were able to be physically present in the classroom and 2) all members within the team agreed to participate in the study. The four distance teams were excluded due to their presence only through technology. Two local teams in which one or more members opted-out were also excluded from this pool.

Because the class met every weekday and a single researcher could observe one team per class period, I chose to observe five teams for this study. I used Microsoft Excel to randomly order the teams within the observation selection pool, and I selected the first five teams (Teams I, K, Q, R and V) on the randomized list.

Measures

In this study, I measured trust, performance and team member behaviors.

Trust. To measure trust, I utilized two instruments that were designed for this study. One survey was administered at the beginning of the semester to assess the initial team trust. Another survey was given weekly to monitor participants' willingness to rely on or "trust" teammates. Each survey instrument is described in detail below.

Initial team trust. To measure participants' willingness to trust before they began working with their assigned team, I created the "Teamwork Process Questionnaire" (see Appendix A) as an adaptation from the Correia, et al. (2008) survey, which itself was based on an instrument by Jarvenpaa et al. (2004). The Teamwork Process Questionnaire has six items with a seven-point Likert scale ranging from strongly disagree to strongly agree (1= Strongly disagree, 4=Neither agree nor disagree, 7=Strongly agree). Then, I conducted a principal

component analysis (with 135 students) to determine if the six trust items on the Pre-survey represented one construct. Only one component had an eigenvalue greater than one and explained 82.4% of the total variance. The scree plot results also showed that only the one component was meaningful. Finally, all component loadings of these six items were greater than .40 (Table 1). Those results indicated that these six items represented one single construct. Reliability analysis of the component showed a Chronbach's Alpha of .95. Thus, these six items could be averaged to represent the concept of "initial trust." A one-way ANOVA demonstrated that there was no significant difference in initial trust levels across teams, $F(23, 112) = 0.77$, $p = .76$, so there were no systematic differences in teams' initial trust levels at the start of the study.

Table 1

Component loadings for the pre-survey trust items

Communalities	**Items**
.95	1. We will have confidence in one another on this team.
.95	2. I will be able to rely on those I work with in this team.
.70	3. There will be a noticeable lack of confidence among those I will work with (reversed)
.94	4. Overall, the people in my team will be very trustworthy.
.92	5. We will be considerate of one another's feelings in this team.
.96	6. The people in my team will be friendly.

Trust throughout the semester. I created the survey "Teamwork Prediction Questionnaire" (see Appendix A) to assess team members' willingness to "trust" each team member, including self, when engaging in their group work throughout the semester. The data from this survey was used to calculate three trust measures: the *trust-by-team* score, the *trust-of-team* score and the *self-trust* score. First I describe the survey and then I explain how the trust measures were calculated.

The "Teamwork Prediction Questionnaire" consisted of a single question: "For each person on your team, including yourself, please indicate how willing you are to rely on this person for a positive contribution to this group activity." The surveys were customized to each team such that they only included the names of that team's members. Participants indicated their level of trust in each individual named on the survey by selecting from a five-point Likert-type scale ranging from low to high trust: 1= Minimally willing to rely on this person (0% to 20%), 2=Marginally willing to rely on this person (21% to 40%), 3=Moderately willing to rely on this person (41% to 60%), 4=Generally willing to rely on this person (61% to 80%), and 5=Extremely willing to rely on this person (81% to 100%). This survey was administered weekly (14 time-points during the semester), thus permitting a longitudinal record of students' trust levels.

The format of the instrument was based on Sweet's (2008) "Prediction Sheet," which was a quick, simple, individualized survey allowing measurement of student perceptions over time. The question for this survey was adapted from the second item on the *Teamwork Process Questionnaire* ("I will be able to rely on those I work with in this team") in order to generate a single question representing student's trust level in another team member. Based on this study's working definition of trust, I worded the item so that it referred to reliance on an individual ("…how willing are you to rely on this person…"). I also added "…for a positive contribution to this group activity" in response to Mayer's advice that a survey instrument that measures trust needs to tap into the trustor's willingness to be vulnerable to the trustee as well as answer the question "trust them to do what?" (Mayer, et al., 1995).

The *Trust by Team* score was a measure of how much an individual was trusted by the other team members. The calculation first required removal of any scores that students assigned themselves on the weekly trust survey. Then, each individual in the class received an

average score from among the scores that participant team-mates gave this individual on the weekly team trust survey. For example, teammate A would have a *Trust by Team* score of 4.6 on a particular weekly trust survey if this teammate received scores of 5, 4, 5, 4, and 5 from teammates B, C, D, E, and F on this survey. Most students had 14 *Trust by Team* scores over the semester. However, if no team members completed the weekly trust survey, the team members received no *Trust by Team* score for that time point. The team level *Trust by Team* score was the mean of the individual members' *Trust by Team* scores. The team level *Trust by* Team score was a general measure of the level of trust within the team.

The *Trust of Team* score was a measure of how much an individual trusted the rest of his or her team (i.e., the team, not including self). This value was calculated by averaging the scores that an individual assigned to the other team members on the weekly trust survey. For example, teammate A would have a *Trust of Team* score of 3 on a particular weekly trust survey if this teammate scored teammates B, C, D, E, and F with the scores 1, 2, 3, 4, and 5. Participants only have *Trust of Team* scores if they completed the weekly trust surveys and scored their teammates.

The *Self-trust* score was a measure of how much an individual was willing to rely on himself or herself. This was the value that an individual assigned to himself or herself on the weekly trust survey.

Performance. The performance measures were then obtained from the graded materials generated within the course. This included the individual quizzes, group quizzes, clicker questions, and peer evaluations.

The *High Member Score* was the highest individual score within the team for any particular quiz or the largest sum of individual quiz scores within a team for a given time frame. The high member score was not always associated with the same individual; different team

members could have the high member score on certain quizzes or within particular time frames. This score was converted to percentage correct by dividing the sum in each time period by the total points possible in that time.

The *Average Member Score* was the average individual score within the team for any particular quiz or the average individual quiz score within a team for a given time frame. This score was converted to percentage correct by dividing the sum in each time period by the total points possible in that time.

The *Group Score* was the team's score for a particular quiz or the sum of group quiz scores within a given time frame. This score was converted to percentage correct by dividing the sum in each time period by the total points possible in that time.

Error checking. Following up on concerns that arose during the study regarding potential inaccuracies in participants' responses on the *Teamwork Prediction Questionnaire*, I created a short, end-of-semester survey called the "Teamwork Survey Post-Analysis" (see Appendix A). The purpose of this survey was to ensure that the data collected from the other surveys during the semester were valid. Participants were asked about how seriously they took the *Teamwork Prediction Questionnaire* and whether or not completing the survey in class while surrounded by teammates influenced their responses.

A one-way ANOVA of the question "Did having team members sitting next to you during class influence your responses on the Teamwork Surveys?" revealed no significant differences across teams, $F(23,66) = 1.27$, $p = 0.23$. A one-way ANOVA of the question "How seriously did you take the Teamwork Survey?" revealed no significant differences across teams, $F(23,66) = 0.70$, $p = 0.83$. Most teams indicated that having team members sitting nearby did not influence their responses, and most teams thought at least a moderate amount about their answer

choices. Thus, the results from *Teamwork Prediction Questionnaires* are probably truthful and accurate.

Researcher as an instrument. Although the researcher description is often omitted from quantitative research, qualitative researchers are careful to identify the researchers' characteristics because "the researcher is the primary instrument for data collection and data analysis" (Merriam, 2002, p. 5). As an instrument, the researcher is an effective tool for collecting and analyzing data, but humans have inherent biases which cannot be eliminated. Qualitative researchers doubt that any people, even researchers, are capable of being truly objective. They instead opt to make researcher biases transparent so that any subjectivities that might shape the researchers' data collection and interpretation can be monitored. Following this lead, I submit my self-description in an attempt to identify my characteristics and potential biases as an instrument.

At the time of this investigation, I, an American woman of European descent, had a Doctor of Veterinary Medicine and had completed my doctoral coursework and preliminary examination in Curriculum and Instructional Technology. At the time I had conducted the study, I had also worked at the same veterinary school within the Office of Curricular and Student Assessment, which was involved with evaluating instructional quality within the institution. Prior to this study, I was not personally familiar with the student participants in this study; however, it was possible that I had analyzed data related to these students as part of my duties in the assessment office. Although my clinical pathology residency training had ended several years before this study began, that role introduced me to several mentors, one of whom was the instructor for this course, and to senior veterinary students I taught in a clinical pathology rotation. My interests in human learning, veterinary medicine, institutional organization and

longevity, group dynamics, and leadership led me to pursue a research topic which combined teamwork, veterinary medicine and learning.

Bias can cause selective attention to details and selective data interpretation. In my case, bias may have resulted from my prior relationships to the instructor and to the field. Memories of my own veterinary school experiences could have helped me better understand the students' perspectives, but it was also possible that unintentionally looking at the participants through the lens of my own experiences and feelings about veterinary training may have distorted the view. To help minimize those concerns, I triangulated my data by utilizing survey responses, course grades, and peer evaluations in addition to comparing my observations with a part-time second observer. Triangulation, one of several common strategies for promoting validity and reliability in qualitative studies, uses "multiple investigators, sources of data, or data collection methods to confirm emerging findings" (Merriam, 2002, p. 31).

Procedures

Survey administration. Participants completed all surveys on SurveyMonkey.com. The *Teamwork Process Questionnaire* and *Teamwork Prediction Questionnaire* were made available during class time, and students completed the surveys using their personal computers. At the beginning of the fourth class period, the instructor provided in-class time for students to complete the *Teamwork Process Questionnaire*. This was the second day team-mates had been together, and the only team activity that teams had performed by that time was the "Grade-Weight-Setting Exercise". Thus, practically speaking, the students completed the survey based on their knowledge of their teammates' identities but without knowing how well these particular teammates would perform within this group. The weekly *Teamwork Prediction Questionnaire* was made available at the end of the week, and students were instructed to

complete the survey after they completed their individual quiz and before they began the group quiz. Most worked on their computers in class to complete the survey; however, some students completed the survey outside of class because the survey remained open for several days. At the end of the semester, students received an email invitation to complete the *Teamwork Survey Post-Analysis* and non-responders received two reminders. Email invitations gave students an opportunity to complete the survey in private, thus encouraging complete honesty. Students who did not complete the survey received email reminders 5 and 12 days later. No class time was provided to complete this survey.

Observations. I observed Teams I, K, Q, R and V once a week for the entire semester. The team observation schedule in Appendix C (page 120) regularly alternated the weekday any particular team was observed in order to minimize any systematic day-of-the-week effects. This schedule was also intended to increase the chances that each team would have been observed for a similar number of quiz days, especially Thursdays during TTQs.

I performed all observations alone or with an occasional assistant observer. During the class, the observer(s) recorded the duration of individual and team quizzes. Whenever team members interacted with each other, we recorded the duration and nature of the interactions. In particular, we attempted to categorize which persons were involved with discussions, who controlled the folder and clicker, who appeared most or least dominant, who appeared most or least talkative, and what decision-making techniques the team used. We also noted any off-task activities by team members, such as playing video games or looking at unrelated websites during class.

During observation sessions, the observer(s) selected seats in the highest row adjacent to the observed team's assigned seats so that the team members were located adjacent to or between the observers. This arrangement maximized the number of team members visible at

once. The observer(s) recorded the each team member's relative seating location to teammates. Although most class sessions were audio-recorded, these recordings were not utilized for analysis due to the inability to identify the speakers.

Immediately after the class, the observer(s) recorded impressions of the class session. If two observers were present, we met and reviewed what we believed occurred during that session. Each observer explained what they saw and discussed how the observations compared until we came to a general sense of what occurred during that class period. These discussions were often audio-recorded. When observing alone, I summarized what I saw in the observation notes. The observation data and participant-generated data were combined to describe emergent patterns, including common features of all observed teams and peculiarities of particular teams.

Compiling data into quarters. The assessments within this study were more frequent and contained fewer questions than other TBL studies about team performance in the literature. For example, Watson et al. examined team and individual performance over three time points using assessments that had between 32-36 questions, and they had few, if any, individual students achieve perfect scores (Watson, et al., 1991). In contrast, this study had at least 28 five-point assessments on which individual students frequently received perfect scores. The low variability at each time point combined with the many time points obscured any potential patterns in trust and performance over time and resisted interpretation by statistical analysis or visual analysis.

Michaelsen et al. (1989, p. 835) provided a solution to the problem because their "data consisted of the cumulative scores obtained from a series of six individual and group tests", each of which "contained 12 to 18 multiple-choice and true/false questions, for a total of 84 to 101 items in all". Taking a cue from their study design, I divided the semester into four roughly

equivalent time periods called “quarters.” Compiling the data into quarters had several advantages. The four distinct time points facilitated more direct comparisons of trust and performance within and between those time points. The increased variability in the data increased the chances of identifying patterns between groups. The data were simplified to permit statistical and visual analysis.

The weekly trust survey data were combined within those periods to produce mean *Trust by Team* scores in each quarter. Similarly, the assessment data was combined within those time periods to generate performance scores within each quarter. The performance scores were calculated by converting Individual and team performance scores performance scores into percentage correct. The following formula was used for individual and team quiz scores respectively:

$$\text{Quiz percent for quarter X} = \frac{\text{Sum of CDQ and TTQ quiz scores during quarter X}}{\text{Total points possible during quarter X}}$$

Analysis

There were three quizzes which were completed only by the team and did not have an individual component. Unless otherwise specified, analyses and reporting of team quiz scores refers to only the team component of those quizzes which had both individual and team components.

Research questions and hypotheses. *Question 1: How does the trust level relate to performance and time? Hypothesis 1a: Trust by team level will predict team performance on group quizzes.* A hierarchical regression analysis similar to that utilized by Watson et al (1991) determined the degree to which the high member scores, average member scores, and *Trust by Team* values contributed to the variability in team scores over the semester. The independent variables high member score, average member score and *Trust by Team* score were entered in

that order. Group scores were entered as the dependent variable. Watson et al. had used a similar order because the literature showed that high member scores, followed distantly by average member scores, were predictive of group results. I utilized team trust as the third independent variable instead of Watson's "group interaction" measure because this study was attempting to describe the relationship of trust to performance. This analysis showed how much the *Trust-by-Team* added to team performance over the best member and average member scores.

Hypothesis 1b: Trust within the team will increase as the amount of time working in the group increased. To investigate the trust over time, I conducted a two-way repeated-measure ANOVA with team number (24 teams) as a between-participant factor and trust-by-team scores (4 quarters) as a within-participant factor. I graphed the patterns of trust by team scores by quarter for each team.

To identify differences in trust between observed teams, a one-way ANOVA compared teams' *trust by team* level within each quarter. The frequency of self-trust and trust of team in observed teams over time were examined for patterns corresponding to the quantitative data.

Question 2: How are team members' trust levels in their team mates related to quiz performance? Hypothesis: Team mates would have higher trust in higher-performing team mates than lower-performing team mates. The "relatedness" of the trust scores was determined by quantitatively and qualitatively comparing observed members' performance with their trust measures. A Pearson correlation was performed within observed teams to compare individuals' mean semester individual quiz score to their trust by team scores. Qualitative analysis involved ranking individuals' trust and performance and categorizing them into "expected trust" and "unexpected trust" based on their performance and trust patterns. The reasoning for "expected" trust was based on the assumption that teammates would be more willing to rely on (trust) the

high performers than the low performing members. Any patterns not adhering to this pattern were counterintuitive and thus were "unexpected" trust. Specifically, expected trust patterns included high performing students receiving high *Trust-by-team* scores or good peer evaluations and lower-performing students receiving lower *trust by team* scores or lower peer evaluations. Unexpected trust patterns involved situations when low-performing students received high trust scores and evaluations, similarly performing students received different trust scores, or high-performing students received low trust scores. Observations and open-ended responses were examined and used illustrate the patterns.

Question 3. What behavior patterns did teams exhibit relative to TBL activities and time? Watson et al. (1991, p. 808) suspected that the interaction patterns in the group ware "a key variable with respect to the quality of the decisions that are reached" and recommended using observer reports instead of or in addition to having participants self-reporting their group interactions. Following this advice, I described patterns in team mate interactions which emerged from researcher observations and participants' comments on open-ended questions. I took special notice of any common patterns between teams' characteristics including decision-making style, whether or not members broke into sub-groups during discussion, and the level and nature of member participation.

The qualitative analysis in this study was based in truth and reality-oriented correspondence theory. This approach presumes that "there is a real world with verifiable patterns that can be observed and predicted" (Patton, 2002, p. 91) and that the purpose of scientific inquiry is to describe reality and determine plausible explanations for verifiable patterns. The merits of this approach include the ability to test claims about effectiveness by using data to examine whether or not the assertions were supported by the evidence (qualitative and quantitative). In the case of this study, were the assertions about the role of

trust in team performance in Team Based Learning supported by the evidence? Postpositivist approaches acknowledge that knowledge, which is inevitably embedded in historical paradigms, is relative and that all research methods are imperfect; thus, using multiple methods (qualitative and quantitative) is important to improve understanding about how the world works and provide evidence to distinguish between more and less plausible claims (Patton, 2002).

Researchers operating from this reality-oriented perspective balance their concerns about validity, reliability and objectivity with knowledge that it is impossible to conduct completely value-free inquiry (Patton, 2002). To address the concerns of personal values and preconceptions which might affect what I saw, heard and recorded, I made explicit my known bias (see p. 47) and confirmed my observations with those of a part-time second observer. To address concerns about validity and reliability, I utilized triangulation of data sources, both qualitative and quantitative, to increase the credibility of the findings.

Handling missing data. Any data generated by the nine students who did not complete the informed consent was deleted prior to analysis. This includes individual scores (individual CDQ quizzes, individual TTQ scores, final exam, peer evaluation score) and any responses on the surveys. Thus, these students did not contribute to the team mates' *Trust by Team* scores. However, because their team mates did generate *Trust by Team* scores about them, these nine students were included when analyzing trust at the team level.

Chapter 4. Results

Question 1: Relationship between trust and performance over time

Preliminary analysis. Over the semester, individuals and teams performed well and often achieved perfect or near perfect scores on their quizzes, so there was a ceiling effect on the scores. Individual and team quizzes had median scores of 81.4% and 97.1% respectively. The trust-by-team levels during this time were also subject to a ceiling effect due to generally high trust scores. Over 90% of the trust-by-team scores ranged between 4 and 5, and the semester median was 4.75. Within teams, teammates generally had high willingness to rely on the other members.

Trust and performance. To investigate whether the trust level predicts academic performance, I conducted a hierarchical regression analysis similar to that utilized by Watson et al (1991). After controlling for the best member and average member scores, there was no predictive relationship between trust and team performance. The predictors were entered in steps 1, 2 and 3: high member score was entered in Step 1, average member score was entered in Step 2, and *trust-by-team* score was entered in Step 3. Group quiz score was the dependent variable. The results showed that only the high member score significantly predicted group quiz scores over the semester (Table 2): In step 1, the high member score positively predicted the group quiz and accounted for 52.5% of the variance in group quiz. In steps 2 and 3, “Average Member Score” and “*Trust-by-team* score” did not significantly predict group quiz and did not account for additional variance beyond "high member score".

Table 2

Hierarchical regression analysis using the high member score, average member score and Trust by Team score to predict group quiz over the semester

Predictor	B	SE B	β	R^2	Adjusted R^2	ΔR^2	Change in F
Step 1 High member scores	0.26	0.06	.73	.53	.50	.53	18.81***
Step 2 Average member scores	0.14	0.13	.32	.56	.50	.03	1.20
Step 3 Trust by Team Scores	0.75	1.23	.11	.57	.48	.01	0.38

Note. * p < .05 **p < .01 ***p < .001

Trust over time. The results of the two- way repeated-measure ANOVA indicated that the main effect of team was significant, $F(23, 369) = 19.71$, $p < .001$; that is, two or more of the 24 teams had different trust-by-team scores over the four quarters. The main effect of quarter was not significant, $F(2.77, 1023.29) = 2.45$, $p=.07$; that means that there was no difference in *trust-by-team* scores over the four quarters for the class as a whole. (Degrees of freedom were corrected using Greenhouse-Geisser because sphericity assumptions were violated.) Finally, the results indicated a significant interaction effect in which team interacted with *trust-by-team* over time, $F(63.78, 1023.29) = 3.07$, $p < .001$. Because the interaction was significant, the interpretation should be based on the interaction effect not the main effects. To investigate the interaction effects, I conducted four additional one-way ANOVA analyses within each quarter. There were two or more teams in each quarter which had significantly different *trust-by-team* scores: quarter 1 ($F(23, 558) = 9.076$, $p < .001$), quarter 2 ($F(23, 578) = 15.11$, $p < .001$), quarter 3 ($F(23, 403) = 12.33$, $p < .001$), quarter 4 ($F(23, 425) = 13.23$, $p < .001$). When the teams' *trust by Team* scores were graphed over time, no universal trust pattern was noted among all teams; however, the patterns could be grouped into several categories based on appearance including

curves which increased, decreased, remained horizontal, were S-shaped, or were U-shaped. Thus, teams experienced a wide variety of trust patterns across time, including increasing trust, decreasing trust, and various fluctuations of trust.

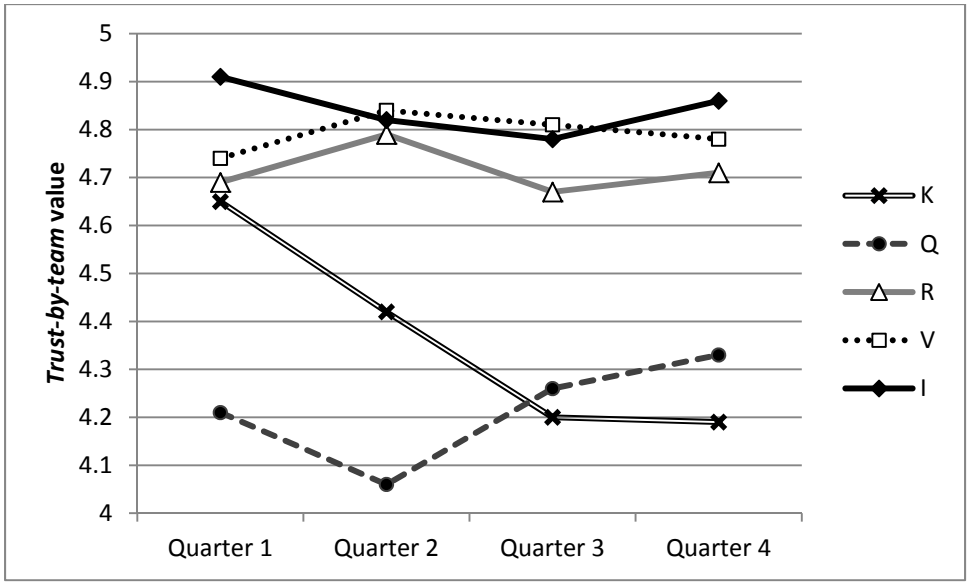


Figure 1. Trust by team within observed teams over time

Because we only had observation data for five teams and we wanted to relate the qualitative findings with the quantitative findings, we re-ran these analyses for only the five observed teams. The results of the two- way repeated-measure ANOVA indicated that the main effect of team was significant, $F(4,83) = 20.50$, $p < .001$; that is, two or more of the five observed teams had different trust-by-team scores over the four quarters. The main effect of quarter was not significant, $F(2.31,191.49) = 1.74$, $p=.17$; that means that there was no difference in *trust-by-team* scores over the four quarters for the five teams. (Degrees of freedom were corrected using Greenhouse-Geisser because sphericity assumptions were violated.) Finally, the results indicated a significant interaction effect in which team interacted with *trust-by-team* over time, $F(9.23, 191.49) = 3.27$, $p < .001$. To investigate the interaction effects, I conducted four

additional one-way ANOVA analyses within each quarter. There were two or more teams in each quarter which had significantly different trust-by-team scores: quarter 1 ($F(4, 118) = 10.82$, $p < .001$), quarter 2 ($F(4, 118) = 17.18$, $p < .001$), quarter 3 ($F(4, 88) = 8.69$, $p < .001$), quarter 4 ($F(4, 89) = 24.66$, $p < .001$). Thus, for the five observed teams together, there was no difference in trust between the quarters; however, there was a difference in trust between two or more teams within each quarter. To determine which teams had significantly different *trust-by-team* scores within each time period, I conducted post-hoc analyses on the significant one-way ANOVA results (Table 3).

Table 3

Means, standard deviations, and post-hoc comparisons of observed teams' trust-by-team values by quarter

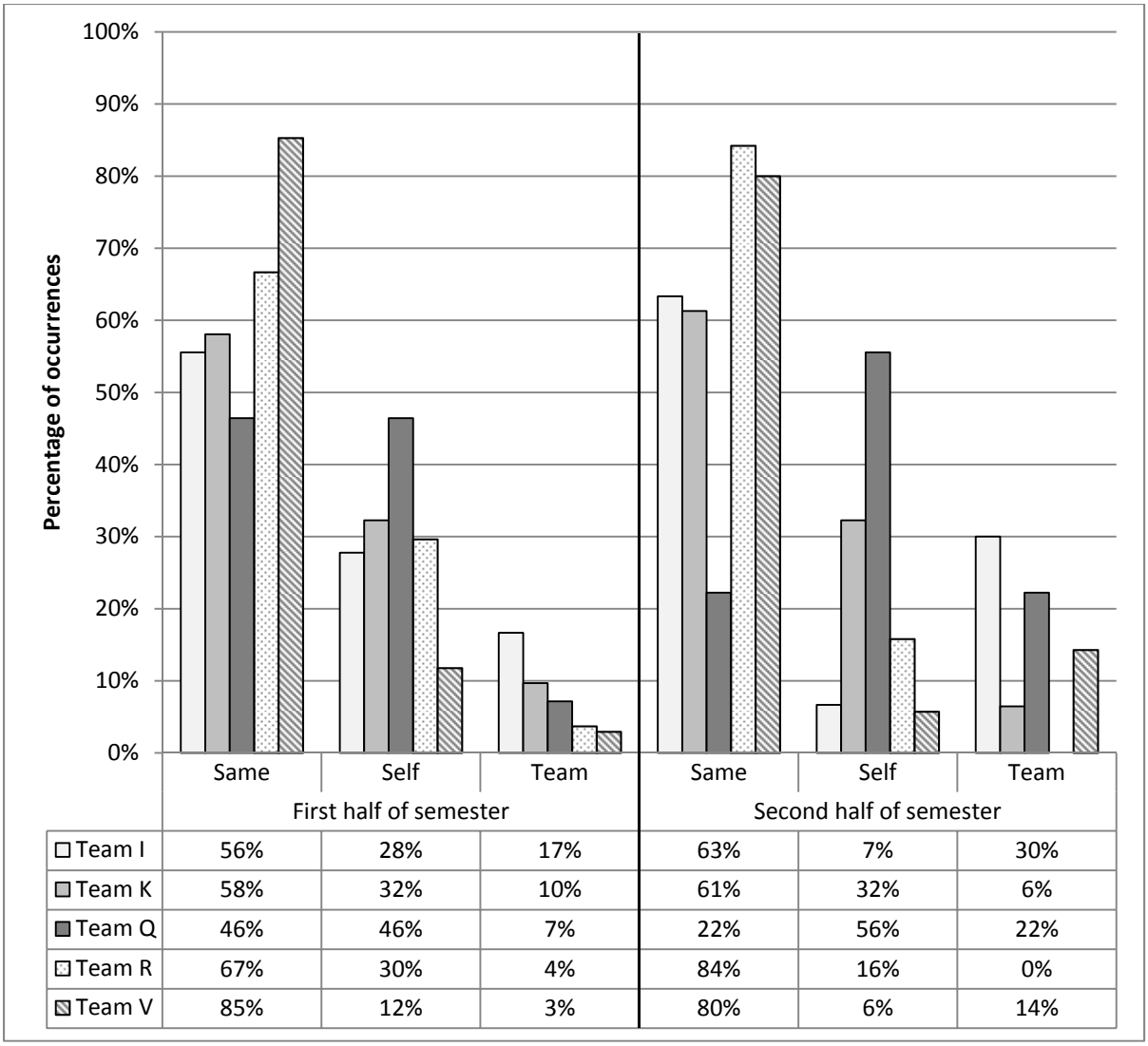
Team	Quarter 1		Quarter 2		Quarter 3		Quarter 4	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
I	4.91	0.16	4.82	0.16	4.78	0.23	4.86	0.20
K	4.65	0.20	4.42	0.38	4.20	0.41	4.19	0.26
Q	4.21	0.58	4.06	0.76	4.26	0.72	4.33	0.40
R	4.69	0.22	4.79	0.36	4.67	0.35	4.71	0.18
V	4.74	0.13	4.84	0.16	4.81	0.18	4.78	0.14
Post-Hoc Comparison*	Q < I,K,R,V		Q < I, R, V; K < I, V		Q < I, V; K < I, R, V		Q < I, R, V; K < I, R, V	

Note. *Only significant comparisons are listed. To reduce type 1 errors, the Bonferroni adjustment was used to identify significant comparisons [$p \leq .05/5 \leq .01$].

Since teams Q and K had significantly lower *trust by team* values than other observed teams at particular times during the semester, I describe some characteristics observed in teams Q and K which distinguished them from the other teams. Teams Q and K differed from teams I, R, and V in their self-selected seating arrangements, social conversation patterns, and self-trust levels.

Social patterns of lower-trusting observed teams. Team K members, unlike those in the other four observed teams, frequently tended to sit in different locations relative to other team members. In the other four teams, the members tended to remain in a certain self-assigned seating arrangement for the entire semester, with the exception of shifting places within the same row. Outside of team K, it was uncommon for members in front and back rows to switch places. The seating arrangements are described in greater detail in Appendix C. During lecture times, team members tended to socialize quietly with the teammates within their same row rather than talking with the persons in front or behind them. Team K members would occasionally chat with persons who happened to be sitting next to them, whether they were team mates or not. One team K member was absent (excused) during half of quarter 2 and all of quarter 3. Team Q was the most rowdy of the teams, and the team members frequently chatted and joked loudly with each other and they often spent much time in extended discussions explaining the answers to each other, even after clicking in the response.

Self-trust and team-trust patterns of observed teams. By comparing *self-trust* and *trust of team* values, individual team members were categorized as having either trusted self more than team ("self"), trusted team more than self ("team"), or trusting both equally ("same") for each survey. According to TBL, team members should become more willing to listen to each other with time because they trusted more; thus, one would expect higher self-trust in the beginning and higher trust of team by the end. This was not true for teams K and Q. Unlike teams I, R, and V, who shifted away from "self" trust towards "same" or "team" trust, teams K and Q did not decrease self-trust between the first and second halves of the semester (Figure 2). Over the semester, no members in teams I, R, and V had greater self-trust than trust-of-team, but half of the members in teams K and Q had higher self-trust than trust-of-team.



	First half of semester			Second half of semester		
	Same	Self	Team	Same	Self	Team
Team I	56%	28%	17%	63%	7%	30%
Team K	58%	32%	10%	61%	32%	6%
Team Q	46%	46%	7%	22%	56%	22%
Team R	67%	30%	4%	84%	16%	0%
Team V	85%	12%	3%	80%	6%	14%

Figure 2. Frequency of self-trust to trust-of-team between first and second halves of the semester

Note. "Same" means the self-trust equals the trust of team. "Self" means self-trust was greater than trust of team. "Team" means trust of team was greater than self-trust.

Question 2: Relatedness of trust and individual performance

The "relatedness" of the trust scores was determined for observed teams by quantitatively and qualitatively comparing members' performance with the trust measures. Patterns in which teammates demonstrated greater willingness to rely on (trust) the higher

performers were categorized as "expected" trust. Any patterns not adhering to this assumption were categorized as "unexpected" trust. Teams I, Q, and R had positive associations (expected trust tendencies) between performance and trust while teams K and V had negative associations (unexpected trust tendencies) between performance and trust. Only for team R did the Pearson correlation indicate a significantly correlated relationship between individual performance and trust in that person (Table 4).

Table 4

Pearson correlations of the semester mean individual quiz scores and Trust by Team within teams

Team	R	N
I	.316	6
K	-.682	6
Q	.712	6
R	.971***	6
V	-.581	6

Note. * p < .05 **p ≤ .01 ***p ≤ .001

In team V, the lowest scoring member had the highest *trust-by-team* score, and the second highest performing member had the lowest *trust by team* score. By trusting a low performer more than a high performer, this team engaged in unexpected trust. The high performer with low trust was talkative, interrupted teammates, and was not able to explain her answers meaningfully, which may have influenced others to "trust" her less. The high performer with low *trust-by-team* "often knows the correct answer but struggles sometimes to express it in a meaningful manner." The low performer with high *trust-by-team*, on the other hand, tended to be quiet. Teammates felt this person "…contributes valuable input during group discussions and readily looks up supporting information during our group discussions of many concepts," "…explains her choices well…," "…has helped focus the group in several discussion situations,"

and "…brings up when she doesn't agree." The low performer with high trust also had lower self-trust scores than teammates which could indicate self-awareness about ability; however, the student also had lower trust of team, which might mean the student was generally less trusting.

As expected in team R, the two persons with the lowest quiz rankings also had the lowest *trust-by-team* scores. The member with the highest quiz rank had the highest *trust-by-team* score. Because the performance corresponded positively with the *trust by team* values, this team generally appeared to have expected trust patterns. The low performers with low trust were observed to be the quieter members of the group, and both had high self-trust, which may indicate inaccurate self-assessment. Team mates complimented one of the low-trust, low-performers on "thinking through the cases and pointing out things that the rest of us didn't catch." They complimented the second for being "very good at explaining why he chooses specific answers" and being "really good at challenging the group on a question if he questions it's [sic] accuracy and many times turned group opinion!" However, the second low-trust, low performer was observed to sometimes be distant and less involved in teamwork and other members encouraged this student to provide more input during discussions: "Could contribute and participate more, speak up! Good contributions when you participate." This student indicated high willingness to rely on self, which could indicate inaccurate self-knowledge or less ability to self-assess reliability under the circumstances. Observers noted that the high performer with high trust dominated the conversation and potentially overwhelmed or alienated some team members; however, positive peer evaluations and feedback suggest this student was actually well-received as a team leader and teacher. Teammates indicated that this high performer did "…a great job as a leader and keeping everyone on track" and was "…really excellent at teaching and breaking things down so other people can understand them." Even

teammates that observers thought may have been overwhelmed by this high performer actually saw this person as a great asset to the team: “She always seems to have the right answer! Is always really prepared and is awesome at making sure everyone's opinion is included in group discussion.” The high performer with high *trust-by-team* actually had lower self-trust and trust-of-team than teammates. It is possible that this individual was generally a less trusting person.

Team Q tended towards expected trust similar to team R, but not to the same degree. The second lowest performer had the lowest trust-by-team. This person was a quiet member who was the “resource” because she looked up information in the notes; however this person also exhibited problem behaviors. Teammates perceived this person to be aloof, defensive, unprepared, non-participatory, and unable to support arguments. Teammate comments about this person included: 1) “…sometimes i [sic] feel you are not "with" us. i [sic] would like to see that change and your participation increase.” 2) “…one suggestion is to try to work on explaining things without getting defensive; passion is a good thing when used appropriatly [sic]. It will make people enjoy your company more!” 3) “…the quiet sometimes causes me to wonder if we're up to speed on all the assignments in a way to contribute. Honestly, I'd prefer that if any of us is behind or confused that they speak up so that as a group we can catch eachother [sic] up andget [sic] everyone to move ahead in the coursework.” This person had high self-trust which could indicate inaccurate self-assessments. The highest performer in the team had the highest trust-by-team score. Teammates called this highly trusted high performer “the teacher” who was “like a fountain of knowledge” because she knew the material, was willing to defend the answer, willing to ask questions and could “usually explain it pretty well when there’s confusion within the group.” One teammate commented on a problem behavior exhibited by this individual: “You are very knowledgeable about clin path but you talk a lot while [the course instructor] is answering people's questions before a quiz. It also seems like

you are not part [of] the group at times." These behaviors, however, must not have been too problematic since this person had the highest trust score. This highly trusted high performer had high self trust and also had the lowest *trust of team* of the members.

Team K had examples of both expected and unexpected trust. One team member exemplified the expectation of high trust corresponding with high performance. The highest performer had the highest peer evaluations and the third highest trust-by-team scores. Teammates positively perceived this highly trusted high performer as a knowledgeable team leader and organizer. Two team members exemplified the unexpected trust pattern of highly trusting low performers: the two lowest performing students also had the two highest trust-by-team scores. Team mates recognized that the lowest performer was a quiet student, but they valued this person's participation, input and thought-provoking questions: "I think you add alot [sic] to the group by asking great questions when you dont [sic] understand something. As important as it is to know the answer, sometimes it is equally as important to find out why that is the answer if you don't understand it. I think we all learn from that." This highly trusted lowest performer had low self-trust which may indicate appropriate self-assessment; however, this person also had the second lowest trust-of-team, which may indicate generally low trust in self and others.

In team K, the three examples of unexpected trust involving high to moderate performers seemed to involve problem behaviors. In the first example, the third best performing student in the team also had the second lowest trust-by-team score and lowest final peer evaluation score. This quiet student was perceived as being generally knowledgeable but was occasionally unprepared: "There have been a few times when you are not caught up on your studying, but other than that its [sic] been good." and "spend a little more time with the material so you can fully understand it". On one survey, this person acknowledged the personal

failing recognized by teammates by assigning a low self-trust score and commenting as follows: “I wasn't really prepared for this quiz the first day we worked on it.” In the second example, teammates felt that this moderate performer with a low midterm peer evaluation was generally knowledgeable but did not “contribute in a positive way” due to inattention during class and missing class. This person was doing “a good job of trying to explain answers,” but a team mate did not appreciate this student’s approach: “many times you explain answers that we all agree on and understand and it takes unnecessary time.” In the third example, the second highest performer with the lowest trust-by-team may have occasionally failed to communicate uncertainty thus appearing more certain than s/he really was: “You are very good at sharing your thoughts which is good. Sometimes, however, there have been times that you aren't quite sure on the material, which is fine, just be sure to let us know when you are unsure.”

Team I provided a mixed picture of expected and unexpected trust. One moderate performer had a much lower trust-by-team score than teammates, but team mates’ comments about this were positive and did not provide clues why this person was less trusted than others. Meanwhile, the lowest performer had the second highest trust-by-team score. This person was quiet, but would contribute positively when choosing to speak: “Often quiet but, contributes meaningfully when needed” and “has been good, could maybe speak up a little more, but always very prepared for the topic” and “The strong silent type, you always knew when to put your 2 cents in.” Both of these team members had lower self-trust than team mates and had greater willingness to rely on teammates than self. Demonstrating expected trust, the two highest performing members had high trust-by-team scores. One of these highly trusted high performers was “quiet, but willing to explain things she felt strongly about” and “contributed consistently every week.” The second highly trusted high performer was recognized for having consistent high performance: “very reliable-thank you!” and “Consistently contributed, knew

the subject matter very well" and "You have a nice blend between easy-going and knowing what's going-on. Great contributions during every quiz! A great asset to our team."

Specific qualities valued by team. Comments from teammates on peer evaluations and trust surveys indicated qualities that teams did and did not appreciate in their members. Team mates encouraged or praised characteristics they liked and pointed out problem behaviors which opposed the desired behavior. Team mates appeared to desire the following qualities in their peers:

1. Consistently, actively contributing, including offering input and not being too passive during discussions
    - "Always an active participant in group conversation!"
    - "Didn't ever have much to say during group activity. Never volunteered to talk in class even though I'm fairly sure she knew what was going on."
    - "Depending on [redacted name] does on her indv directly results in how much she participates in the group quiz. If she feels like she failed the indv quiz she basically gives up for the group on and does not participate."
2. Being intelligent, knowledgeable, and having the correct answers
    - "intellingent [sic]. knows subject matter."
    - "She always seems to have the right answer!"
    - "You usually know your material and cases pretty well and offer sound advice."
3. Listening to team mates, including ensuring all have had the opportunity to contribute before submitting answers
    - "always willing to listen to others and consider their reasoning"
    - "group works fairly well together, we typically are open to other's opinions when we have differences on the quizzes"

- "Sometimes it seems like you move on ahead without consulting the group, maybe wait so ensure everyone has had a change to comment"

4. Teaching the other team mates about the content
    - "One thing that I appreciate about your role in our group is you tend to be really excellent at teaching and breaking things down so other people can understand them. "
    - "Sometimes more quiet than I'd like, but alway [sic] brings something to the table and is able to elucidate some of the more technical points in ways that I think help us al [sic] to better understand the topics."
5. Confidently stating answers and standing up for one's answers when differing from the group
    - "… always contributes to group brings up when she doesn't agree"
    - "You always do a good job of brining [sic] in a different perspective or think about it in a different way"
    - "Always on top of things an [sic?] not afraid to be wrong and commit to the answer anyway. It's a far better thing to try to advance one's ideas so that we know that we're all participating and trying. The best way for all of us to learn is to kick around both the right and wrong trains of thought so we can learn what's correcct [sic] and what traps lie ahead."
    - "you should be more confident in your opinions on quiz answers and help sway the group!"
6. Having pleasant attitudes, including being positive and supportive, providing some comic relief, being team-oriented and not being defensive
    - "Willing to be a team player and work through problems together"

- "You know how to take the business seriously without taking yourself too seriously."
- "Often confused about the concepts, could be fairly argumenitive [sic] or put off with disagreaments [sic], often appeared distracted by facebook."

7. Knowing when to speak up
    - "The strong silent type, you always knew when to put your 2 cents in. Overall great contributer [sic] to our team."
    - "Often quiet but, contributes meaningfully when needed."
8. Being prepared for group activities
    - "you do a good job of having read to know"
    - "has been great, always prepared, states her reasoning well"
9. Asking thought-provoking questions and challenging members to clarify their thoughts or reasoning
    - "[Redacted name] is good at getting group members to justify their answers. This has helped us all learn the details of different concepts as a group"
    - "He's really good at challenging the group on a question if he questions it's [sic] accuracy and many times turned group opinion!"
10. Asking the team for help when confused
    - "He is always helpful during group discussions and if he doesn't understand something he asks for help from the group."
    - "Always prepared and willing to share her thoughts, even to the point of simply saying, "you know, I just don't get this, could someone explain it?" I think that's a valuable quality to have, certainly will advance her learning in the long run."
11. Willingly admitting when uncertain about an answer

- "You are very good at sharing your thoughts which is good. Sometimes, however, there have been times that you aren't quite sure on the material, which is fine, just be sure to let us know when you are unsure."

12. Providing clear supporting evidence for answer choices, presenting answer choices succinctly and persuasively, and thoroughly analyzing the topic
    - "… makes her opinion known and clearly can explain why she picked what she did."
    - "You did a nice job of thinking through the cases and pointing out things that the rest of us didn't catch"
    - "… often knows the correct answer but struggles sometimes to express it in a meaningful manner"
13. Facilitating the group function by keeping the team on task or performing well a role within the team
    - "… you were always willing to write an appeal if one was needed. "
    - "You did a great job of looking things up in the book!"
    - "You are an awesome clicker person and will be greatly missed!"
    - "… you were our "group quiz submission person" and you were very good at it."
    - "You do a great job as a leader and keeping everyone on track."
    - "Always includes everyone in discussion"
14. Present physically and mentally, including remaining attentive, on-task, punctual, and non-disruptive
    - "Seems like she has missed more classes than other group members."

- "You are very knowledgeable about clin path but you talk a lot while [the teacher] is answering people's questions before a quiz. It also seems like you are not part the group at times."
- "Often you are not focusing on class but rather on facebook or working on upcoming Pathfinder cases; this make it difficult to discuss topics with you then since you aren't "uptodate" on what the discussion is"

Although teammates would not tolerate an individual who was unprepared, they appreciated good questions from confused teammates and preferred that teammates inform the team when unsure about an answer. A low-performing student could be perceived as contributing positively if this person spoke up only when certain about the answer or if the person asked thought-provoking questions which helped the team learn. On the other hand, the team may have valued less the contributions from a high performing students who occasionally provided an uncertain answer without warning the team, who were sometimes unprepared, or who were unable to convince teammates to choose their correct answer.

Attempts to map these desired behaviors against the group interaction framework from Watson and Michaelsen (1988) were unfruitful for further evaluating if these behaviors facilitated or interfered with group performance and were thus reasonable expectations. First, the desired behaviors identified in this study did not sufficiently match the behaviors identified in the framework to make a viable comparison. Second, the framework used concepts, such as "cohesiveness" which were comprised of multiple behaviors and some behaviors were shared among multiple concepts. This overlap, made it difficult to determine the positive or negative influence of particular behaviors on group function.

Question 3: Team behavior patterns

Teams I, K, Q, R and V differed in style and their members interacted with each other in different ways. Each team is described below.

Team Q was the lowest performing team. This team appeared the most chaotic, and members were loud and passionate. They frequently broke into subgroups, but usually worked towards consensus. It was uncertain how well team members listened to each other. The use of simultaneous subgroups and interruption during discussion became so bad at one point that one team member called a time out and indicated that everyone was talking at once and she was unable to follow the discussion.

Team V was the best performing team and the second-highest trusting team. One notable characteristic about this team was the low volume and long, thoughtful pauses in the teammates' discussions. This team rarely used subgroups during CDQ and TTQ discussions and was never observed voting. This team would listen to each other and discuss disagreements until all members were in agreement. Even when working fast, the members took turns and avoided interruption, only one individual tended to occasionally interrupt. Team V had a significantly higher individual quiz mean for the semester than team Q, $t(54) = 3.37$, $p = .001$, and team R, $t(54) = 2.28$, $p = .027$.

Team I generally exhibited mid-range performance. They exhibited discussion behaviors which were mid-way between the extremes exhibited by Teams Q and V. For example, during discussions the students would occasionally break into subgroups which talked simultaneously, but then they would come back together to discuss further. Sometimes the team appeared to have agreed on an answer but discussed anyway, perhaps to ensure understanding. They were the highest trusting team observed.

Team R appeared to have one very dominant member. Early in the semester, some team members sat more physically distant from the team than other teams. Several teammates also seemed quiet and socially distant. Later, though, members sat closer together and seemed to engage more in discussions.

Team K members frequently sat in different seats each class period. They worked in subgroups less frequently than team I and seemed to take turns, speak slowly, avoid interruptions, listen to each other, explain things to each other and build on each other's ideas. This team had one member who had an excused absence from class for about 1.5 months. This student was absent during half of quarter 2 and all of quarter 3.

Team decision-making and voting. The 19 teams (with only opt-in members) participating in the study completed 23 Case Discussion Quiz pairs (individual and group components) with 5 questions each for a total of 2185 questions. Of these, the students in class incorrectly answered 82 questions on the team quizzes (4% of questions on team quiz were answered incorrectly). Generally, teams answered questions incorrectly more frequently when a greater proportion of individuals incorrectly answered those questions (Figure 3). Incorrect team answers were most frequently submitted by teams that had 32% to 17% correct members (mostly one person correct), but wrong team answers were submitted less frequently when 16% to 0% of members were correct (mostly no correct members). No teams missed questions on their group quiz when 100% to 84% of individuals had correctly answered a question. For 80% of the incorrectly answered team questions, teams had selected the most frequently held (majority) answer in the team.

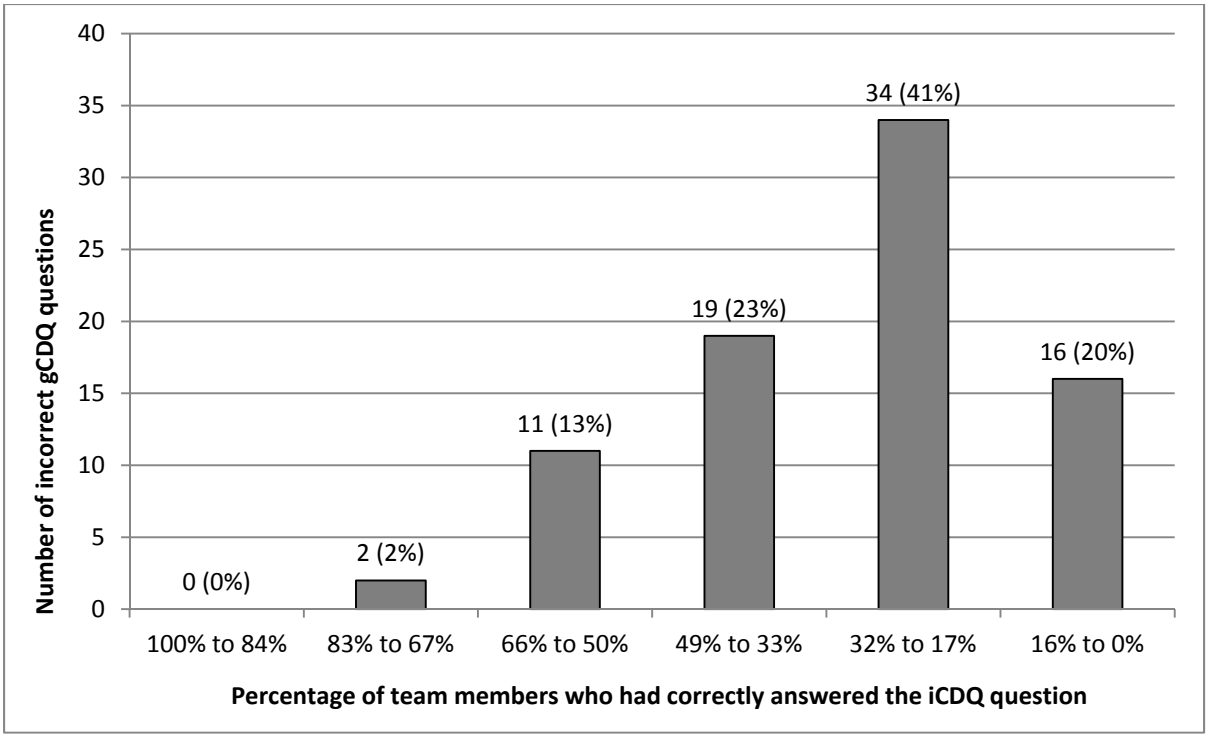


Figure 3. The frequency of incorrect team answers provided by teams with a certain percentage of correct individuals

The situations which were most likely to provide interesting information about group processes and decision-making included the following:

(a) all members were correct, but the team was incorrect.

(b) all members were incorrect, but the team was correct,

(c) a single person was correct, but team was incorrect,

(d) a single individual was correct, and the team was correct,

(e) one person had the same incorrect answer as the team.

To evaluate these, 60 iCDQ questions which had only one correct team member or no correct team members were added to the analysis. Regarding the first situation, there were no occasions when all team members were correct, but the team was wrong. Regarding the second situation, there was only one instance in which the team answered correctly when all

teammates answered the question incorrectly (Table 5). In that case, Team Q correctly chose the answer "a.There is an increased demand for phagocytes in the tissues," but five team mates chose "e.The lymphocytosis is attributed to a stimulation of lymphopoiesis in the bone marrow" and one chose "c.The bone marrow neutrophil storage pool is partially depleted." Because this event occurred when the researchers were observing a different team, no observation data is available to further illuminate the group process. Regarding the third and fourth situations, the team answered correctly more often with one correct member present than with no correct members. Yet, even with one correct member, the team frequently chose the wrong answer, so correct individuals appeared to have difficulty persuading the group. Regarding the fifth situation, the team had the same answer as a single incorrect person for 7% of the 82 questions missed by the team (Figure 4). This suggests that single incorrect individuals could rarely convince the team to choose their position. On three of the six occasions when the team selected the same wrong answer as a single incorrect individual, there was also a single correct individual. This occurred twice with team W and once with team I. Unfortunately, no observation data was available for these occurrences to help determine how one individual was able to prevail over another individual.

Table 5

Frequency of correct and incorrect team answers when there were 0 or 1 correct individuals

Number of Individuals Answering Correctly	**Number of Team Answers**		**Total**
	Correct	**Incorrect**	
None	1 (6%)	15 (94%)	16
One	12 [27%]	32 [73%]	44
Total	13	47	60

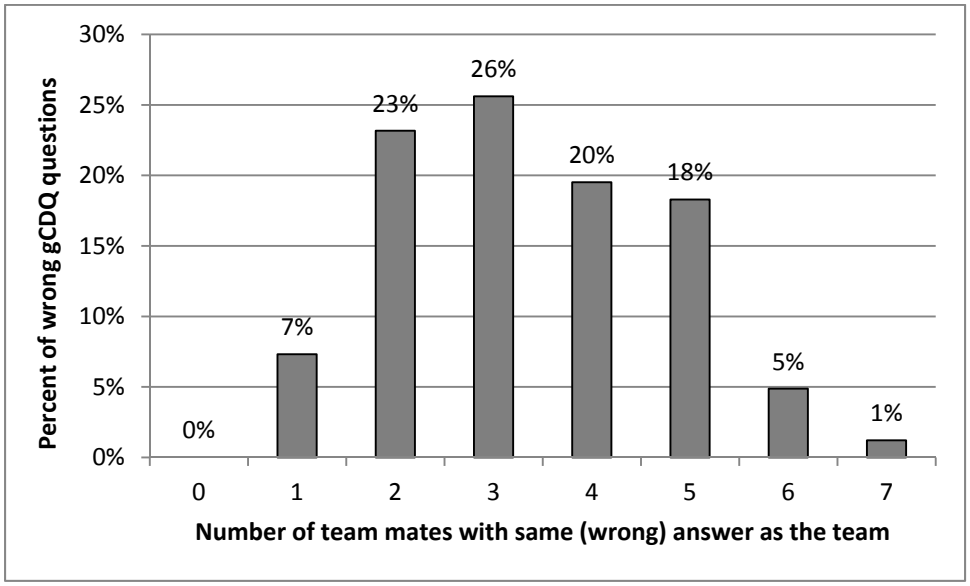


Figure 4. Percent of instances in which a certain number of individuals had same incorrect answer as team

Voting was observed on three occasions during quiz discussions, and the team chose the incorrect answer each time. On one occasion, team K unsuccessfully used majority voting to choose an answer. In that case, one student had a perfect score and was the only team member with the correct answer for the question missed by the team. It would seem this student was unsuccessful in convincing the others to choose the correct answer. Team Q used majority rule on two occasions. On one occasion, each sub-group had a different answer and they could not agree. One member suggested they choose one and if it was wrong, then the subgroup proposing the wrong answer must write the appeal. The team took a vote, but they ended up choosing the wrong answer. On another question, the team again was unable to resolve their disagreements, so a team member suggested a vote. One dissenting person abstained from voting and another dissenter gave in. The team selected the incorrect answer, yet one dissenter had the correct answer for that question.

Task-related patterns. Discussion patterns varied depending on whether the group discussed during clicker questions or RATs with individual and team components.

Activities with both individual and team components. The teams usually had similar approaches for the team component of CDQs and TTQs. Students spent about 4.6 minutes per CDQ question (3.6 minutes per question on the individual portion and 1.0 minute per question on the team portion). They spent about 8.1 minutes per TTQ question (7 minutes per question on individual and 1.1 minutes per question on team portion). During the discussion after the individual component, the front row team-mates usually would turn their seats to face team-mates in the back row. Then one team member read each quiz question one-by-one and team members would offer their answers after each question number was announced. If the team-mates agreed, the team would move to the next question. If someone disagreed, the team-mates would give their reasoning for the answer and try to convince each other by explaining the rationale for the answer they support. Generally team-mates took turns speaking, listened to each other, avoided interrupting and worked towards consensus decision-making, especially in teams I, K, and V. All except team V broke into subgroups with varying frequency during discussions and then came back together to make the final decision. During the discussion, team members may also have asked each other questions and explained concepts to each other. Team members appeared to clearly discuss all the quiz questions prior to selecting the final answers.

Clicker question discussions. Teams had several styles when discussing the quick clicker questions that occurred during the lecture. Clicker question discussions tended to be very fast, and students answered clicker questions at a rate of one question per minute on average. At times, team members appeared to communicate using body language (i.e. glances, nods, and hand gestures). Generally the discussion process differed from discussing the quizzes

following the individual component. Unlike the quiz discussions, the front team members would often keep their chair facing forward; however, they sometimes turned their heads or partially turned their bodies if they wanted to communicate something to the back row. The discussion and clicking styles included: 1) generating whole-group consensus answers in an orderly manner with most team-mates participating by contributing their answers or approving the suggested answers, 2) dividing the team into sub-groups which may or may not come to obvious whole-group consensus prior to clicking, 3) unilateral clicking when the person with the clicker submits (clicks) the group answer without obviously consulting with other team-mates, or 4) minimal consult clicking when the clicker person clicks the team answer after only obviously consulting with a portion of the team, whether a subgroup or a specific individual. Some teams had the same person using the clicker week after week and others had more people utilizing the clicker.

Teams sometimes transitioned between various discussion techniques for different questions within a series or while discussing a particular question. The four techniques could be found in the observed teams, but some teams may have utilized certain techniques more frequently than others. Forming subgroups was common, and subgroups usually included team members sitting in the same row. Sub-groups often held separate discussions simultaneously, and then the subgroups sometimes came back together to share the conclusions determined in each subgroup.

Unilateral or minimal consultation clicking produced varied reactions in teammates. In many cases, the team-mates did not appear to be agitated or voice objections at times when answers were clicked in by one team mate or sub-group without obvious consultation with the others. In fact, on one occasion teammates praised the person with the clicker for the quick clicking, and one student was frequently observed to click in without obvious discussion, yet

several teammates complimented this person's clicker work. Tacit approval may have been occurring in some cases, as illustrated by one teammate's written words to another: "Sometimes it seemed that you were distant since you didn't always turn around to give input - but I think it was more that you "trusted" us and heard what everyone was saying and agreed."

However, the unilateral or minimal consultation clicking was upsetting to some teammates. One student wrote: "Things are generally going well with my group, but there are 1 or 2 people who are not inclined to participate. When they have the clickers, they click the answer without doing any discussing, and I have to ask them which answer choice they chose! It's a little discouraging..." Another student wrote to a peer on the evaluation: "I appreciate your willingness to discuss things and explain why answers are right or wrong. It would be nice to start doing this before we click in our group answer." Another person was upset when the clicker person only consulted part of the team: "It seems like the person who has the clicker discusses the question only with those sitting immediately to their right or left without discussing with the rest of the group. By the time I read through the question on the screen, the answer has already been clicked in. It's a bit frustrating." This problem occurred in team Q because one team-mate commented, "Many times you will turn to one member in the group, ask what they think the answer was, and then click in the answer without asking the group. " Sometimes the problem appeared related to being too quick to click in before consensus was achieved. In team R, a team mate liked how S109 would "refer back to the group on nearly every question" when in charge of the clicker, but the teammate also suggested that S109 "slow down a bit when clicking in or getting the answer from the group until everyone has heard what the final say." Similarly, in team Q, S098 was once observed to have changed the answer several times because she had clicked in the middle of discussion before consensus had been achieved.

Participation with time. Some groups moved to greater member participation and improved team interactions with time. Early in the semester (three weeks of teamwork), at least one student expressed mixed feeling for teamwork: "I like the team idea; however, I am starting to realize that I might be stronger as an individual. The team dose [sic] provide sound judgment on group quizzes, and we really do work great together." Although it is uncertain whether this particular person's mind changed with time, other classmates clearly indicated that groups were improving by the fifth week. A student in team P wrote: "group dynamic is getting better... more unity during decisions. Everyone seems to be bringing something to the table." After seven weeks of teamwork, one student indicated "our team work is becoming more effective as the semester progresses." Some final peer evaluations indicated that certain team members seemed to have improved since the midterm peer evaluation. For example, team members commented about other members' improvement: 1) "After the 'practice eval.' she has contributed more to the group discussions and helps with explaining concepts," 2) "After our first 'practice eval.' he began talking/contributing more to the group discussions," 3) "huge improvement in contribution," and 4) "I noticed an improvement in your participation in the group for the 2nd half of the semester!". After eight weeks working together, a team T member wrote: "Our group has done a great job participating together. I think even some of the shyer (can't spell) people are starting to open up and discuss more. Our group ROCKS!" After about 13 weeks together, a team R member also felt that the team was improving: "whole group really been stepping up."

Although the comments suggested a general trend towards greater participation and better group attitudes with time, a few comments suggest that some team mates did not improve or regressed. After 13 weeks together, one student indicated that a particular team mate "mostly acts like all qroup [sic] activities are a waste of time." On the final peer evaluation,

one student wrote to another: “Since the second half of the semester, you don't seem to be apart [sic] of the group anymore.” Another commented that “some people are speaking up at the end of the semester and others are getting quieter”.

Communication speed, density, and mode. With time, the observers had increasing difficulty following some teams’ interactions. From the observer’s outside perspective, the discussion seemed fast and chaotic; yet, the team members seemed to understand each other well. For example, by the fourth month, the team V discussions seemed to go so fast that it almost sounded like everyone was talking at once but they were not. The discussions seemed information dense because the team squeezed in a lot of interaction within a short period of time. This was particularly remarkable for team V because this team tended to have long pauses between team mates’ comments earlier in the semester. The observers felt the members were increasingly informal and relaxed with each other. One observer reported feeling as if the team members were almost reading each other’s minds when seemingly disorganized discussions unexpectedly concluded in agreement. Sometimes a discussion was very difficult to follow but suddenly a team member would ask “so we’re all agreed?” and the members approved. Sometimes the observers were puzzled when the team reached conclusions that seemed abrupt and did not obviously flow from the observed interaction. Teams increasingly used non-verbal signals and very brief verbalizations to communicate among team members. The observers noted teams using head nods, glances, and hand gestures but may have missed more subtle non-verbal cues that comprised communication among teammates.

Self-assigned roles in the teams. In addition to self-selecting certain quasi-permanent locations to sit within the team, individuals in teams tended to settle into certain roles based on one or more duties they tended to perform. The duties included: (a) logging into WebCT and recording the team responses, (b) reading the quiz questions aloud, (c) keeping the team on

time or on topic, (d) picking up the team folder, (e) operating the clicker (clicker operator), (f) clarifying information, offering varying perspectives, or asking questions to ensure understanding (devil's advocate or questioner), (g) explaining concepts to teammates (teacher), (h) checking facts with literature (resource person), (i) encouraging all to participate, and (j) comic relief. Some duties were more consistently performed by particular individuals in the teams, and these duties became identified as part of the member's roles in the team. The distinctions between roles and the duties for each role varied among teams. For example, in team Q each person seemed to take on a clear role based on their duty, but team V did not appear to have a clear leader and the duties were mostly shared among two or more members. In several teams (I, K, Q), the role of "leader" tended to be the person who facilitated or organized the team discussions during quizzes. In teams I and K, the "leader" also tended to be the clicker operator. The clicker operator, however, was not always perceived as the leader, particularly in team Q where one person was identified as the clicker person and another as the quiz leader.

Chapter 5. Discussion

Relationship between trust and performance over time

In principle, Team-Based Learning is supposed to promote positive group interactions which promote trust, and increasing trust allows members to more freely share their viewpoints and challenge each other during discussions. The more free exchange of information then is supposed to result in better performance and learning. This study, however, showed no predictive value of trust on team performance over the semester. As with Watson et al. (1991), high member score was the most predictive measure of team performance.

According to Watson, et al. (1991), team performance and effectiveness can improve with time when using TBL. If the performance increases in TBL are due to how TBL promotes trust among members, then one would expect trust to increase over time in proportion to the performance. Furthermore, one would expect a consistent trust pattern from all teams experiencing the same TBL course environment if the teaching method was a primary influence on trust levels. However, there was no significant change in trust over time for the class as a whole and no uniform trust pattern among all teams. Some teams' trust increased with time while others decreased, fluctuated, or remained stable. This may reflect a complex relationship between time and group performance. The following observations made by Arrow et al about groups in general, may apply to TBL teams as well. On a system level, groups "appear to change over time—but not necessarily in the same way, or at the same rate" and within groups the "temporal patterns in group interaction and task activity appear to differ across tasks, across groups, and also across apparently similar groups doing identical tasks" (Arrow, et al., 2004, p. 93). A coherent story explaining how and why the various exceptions and contingencies occur has yet to emerge (Arrow, et al., 2004).

Unlike the other observed teams' trust patterns, which remained constant over time, Team Q's trust dropped during the second quarter then increased, and team K's trust dropped over time. The cause for these patterns was unclear, but the literature and my observation provide several plausible theories.

First, the TBL literature suggests that self-orientation can interact with team-orientation and performance in a complex, circular relationship: individuals influence the team performance and the team performance affects individual performance. Effective teams must balance individual needs with team-orientation. Individual differences and self-orientation can benefit teams by adding creativity and leading to questioning of the group processes; however, it can also inhibit teamwork when out-of-balance with team-oriented behaviors (Watson, Johnson, Kumar, et al., 1998; Watson, Johnson, & Merritt, 1998). In this case, the two lower-trust teams exhibited higher self-trust than the other teams, which may have indicated this lack of balance.

Second, both lower-trust teams' behaviors and circumstances might have indicated different team development than the other teams. According to Tuckman's model (Tuckman, 1965; Tuckman & Jensen, 1977) and Wheelan's Integrative Model (Wheelan, 2005; Wheelan, et al., 2003) groups generally pass through a series of distinctive phases (e.g., "forming", "storming", "norming", "performing" and "adjourning") in a linear fashion. However, in some cases disruptions could cause groups to temporarily regress to an earlier, less productive stage, or teams may get stuck in an earlier stage and never develop to full maturity (Sweet & Michaelsen, 2007; Wheelan, 2005), though this outcome is not inevitable (Foldy & Buckley, 2010). Team K's drop in trust during quarter 2 partially coincided with the beginning of one member's 1.5 month excused absence, which may have disrupted the group development. Also, team K had the most fluid self-selected seating arrangements, possibly suggesting flawed group

bonding such that members had individualistic attitudes or ambivalence towards group members. While Team Q did not experience an interruption in group membership, their chaotic and boisterous behavior, coupled with poor performance, suggests that they did not reach the highest stage of group development. This study did not have sufficient data to definitively support these explanations.

In summary, trust, as measured in this study, does not appear to differentiate between higher and lower performing groups within TBL, weakening the assertion that highly-effective TBL "teams" are characterized by high level of trust (Michaelsen, et al., 2004, p. 12). Furthermore, TBL did not produce uniform behaviors or patterns within all teams.

Relatedness of trust and individual performance

As one antecedent of trust, the trustee's competence (ability) is a characteristic that makes that entity worthy of risk (Butler, 1991, 1992; Mayer, et al., 1995; Tschannen-Moran & Hoy, 2000). In an academic environment, grades (quiz scores) are the currency by which individuals' learning success is measured, so high performance should be a prized commodity. Furthermore, the high-performance teams produced by TBL, which are presumably characterized by "a high level of trust among members of the group" (Fink, 2004, p. 12) should be more willingly to rely on (trust) the contributions from higher performing team mates. However I found instances in which other factors seemed to influence trust at least as much as academic performance. This was true both for those who were trusted more or less than expected. More than reflecting absolute performance, an individual's *trust-by-team* scores appeared related to that person's ability to help the team choose the correct answer and understand the material.

The participants falling into the "unexpected trust" category were represented by two exemplars: 1) higher-performing students with low trust and 2) lower-performing students

with high trust. In the case of students who performed well but were trusted less than expected, teammates appeared to recognize these students' intelligence and ability. However, these high-performers also exhibited problem behaviors such as interrupting team mates, failing to notify the team when uncertain about answers, coming unprepared, and being unable to articulate reasons for their correct answers. Such teammates may have been perceived as unpredictable, misleading and less reliable. Students who performed less well, but were trusted more than expected, tended to be quiet and had relatively lower self-trust than their teammates. Based on team mates' comments, such students may have known when to speak up with useful or correct information and when to remain silent. Proportionally speaking, a poor performer who spoke only when certain might have had a greater ratio of correct to incorrect responses than a high-performer speaking indiscriminately. Teammates appear more likely to trust these self-aware teammates than other quiet, lower performers who had high self-trust despite their relatively lower performance.

The exemplars in the preceding paragraph demonstrate that team mates are reacting to factors in addition to performance when they trust their teammates. These factors were identified within students' comments about what they did or did not appreciate in team members (see page 66), and coincided with the factors of trustworthiness identified in the trust literature: consistency, competence, receptivity, openness, loyalty, fairness, benevolence, discreetness, integrity, and availability (Butler, 1991, 1992; Tschannen-Moran & Hoy, 2000). Student behaviors relate to the trustworthiness factors as demonstrated by these comments by teammates. One unobserved team member, who had mid-range quiz scores, low peer evaluations, and low trust-by-team scores, was accused of throwing the team under the bus [indiscreet] and sitting closer to her friend than the team [disloyalty], tardiness and absenteeism [unavailability], and persuading the group to select wrong answers

[incompetence]. Another individual with the lowest individual quiz scores was trusted more and praised for caring what everyone had to say [receptivity], keeping the group on task [competence], and giving everyone a chance to speak [fairness]. One participant gave a certain team mate a high scores "…not because she gives positive input [incompetence], but because she never [consistency] comprehends the concept at hand and always [consistency] get the individual questions wrong [incompetence], so I learn more from trying to explain my reasoning for the group quiz. So in a backwards sense, she gives me a good learning environment, but doesn't contribute a darn thing herself."

As illustrated by these examples, because trustworthiness has multiple components, it is possible that behaviors which breach trust on certain components could overshadow the individual's positive qualities, making that person appear less trustworthy overall. Likewise, persons might compensate for deficiencies by exhibiting exemplary behavior on certain trustworthiness components. However, determining how team members' behaviors within each trustworthiness factor influenced the group processes in TBL teams was beyond the scope of this study. Relating the trustworthiness of individuals' behaviors with effective TBL group processes could be a valuable area for future research.

Performance may not have had as strong an influence on trust as expected in this study because team members did not know for certain what individuals scored on the quiz they had just taken. When Sweet (2008) found that students in team based testing situations more accurately assessed each other's levels of expertise over time, his experimental design had used IF-AT forms. Also, unlike Michaelsen's recommendations to write individuals' scores in the team folder (2004b; Michaelsen, et al., 2004), this study did not make students' performance scores public within the team. The IF-AT forms and the reporting of individuals' scores could help students accurately identify members with the most correct information. Perhaps when

teammates have greater room for uncertainty when identifying the correct member, the various trustworthiness factors described above can have greater influence on how willing an individual is to rely on (trust) another. Whereas, the same individual may be more willing to overlook problem behaviors if they could be more certain that the teammate was highly likely to be correct. Additional research can explore whether using or not using IF-AT forms and public individual performance reporting can affect how accurately individuals students predict another teammate's performance and how an individuals' confidence level in another's performance influences their willingness to rely on (trust) that person when that person exhibits (or lacks) various other trustworthiness factors other than competence.

Team behavior patterns

The interactions within the team appear to be more complex than initially anticipated. However, team observations and student reports revealed patterns in team members' behaviors, some of which could have implications for learning. With increasing time, group participation and attitudes towards group work improved, and team communication seemed to become increasingly, fast, dense and non-verbal. Although teams tended to develop self-assigned roles within the teams, they did so somewhat differently than each other. The learning implications for these finding are uncertain, and but they do indicate that student attitudes and team processes do change with time and that not all teams within the same teaching system organize themselves in the same way. The following teams patterns related to task type and decision-making had more obvious significance for learning and group performance.

Convincing the team. Logic dictates two likely causes for missed questions on group quizzes: (a) the team did not have the intra-group resources to correctly answer the question (e.g., no individuals had correctly answered the question) or (b) correct individuals were unsuccessful in convincing the team. For 18% of the team wrong answers, all members were

incorrect on the individual quiz; thus, it appeared that only a small portion of the team error was due to insufficient resources. The majority of team error likely came from process loss during team interactions. Teams tended to use a process of determining team responses by polling members, so the answer choices already existing among the members probably became the pool of potential team answers. Predictably, teams did miss all but one of question incorrectly answered by all members. That one exception may have exemplified the process gains possible when the group interaction generates new ideas, insights, or strategies not previously held by any individual member (Hill, 1982; Watson, et al., 2008; Watson, Johnson, & Merritt, 1998). Hackman and Morris (1975, p. 62) suggest that "group performance is proximally controlled by three general 'summary variables': a) the effort brought to bear on the task by group members; b) the task performance strategies used by group members in carrying out the task; and c) the knowledge and skills of group members which are effectively brought to bear on the task". Perhaps in this exceptional case the team utilized superior task performance strategies to overcome the lack of knowledge. Unfortunately, there was no observation data to give us more information about the nature of intra-team interactions.

It was possible for teams to achieve higher scores than any individual member because individuals' incorrect responses for any particular question tended to be staggered within the group. There was at least one correct member for 99% of the questions. Since there were correct answers available among team members, the team's success depended upon its ability to identify which member had the correct information and select that response. How successfully a team could do this seemed to depend on the number of team mates who had the correct answer. The highest frequency of wrong team answers occurred when there was primarily one correct member in the team. Presumably this reflects the difficulty a single *correct* person has persuading the group. It also appeared that a single *incorrect* person might

also have difficulty persuading the team since 7% of the team wrong answers had been held by only one person.

While not allowing an incorrect individual to sway the group can be beneficial to teams, teams can experience process loss if an incorrect majority overrules a correct minority. For this reason, majority voting, rather than consensus, can be a risky decision-making method. Voting fails to produce the correct answer when the minority is correct. This occurred during the three observed instances of voting. Each time an incorrect majority overruled the correct minority to the detriment of the entire team.

Minorities within groups had varying degrees of success convincing their team mates to select their answer. Who is at fault when the team gets the wrong answer due to not accepting the correct answer from a minority or when they accept the incorrect answer from the minority? Students' comments appear to hold the individual or minority responsible for the team's failure. One student implied that a correct individual's unwillingness or inability to convince the others led to process loss: "There have been times when 5 of us were wrong, you were right, but you didn't explain your thoughts to us to let us see we were wrong." In another case, teammates perceived an individual as having power to (negatively) influence the group decisions: "When she is here she doesn't know what is going on & asks dumb questions & persuades us sometimes to pick wrong answers!" One way a correct individual could sway the team appeared related to that person's ability to clearly communicate a strong rationale for an answer: a) "Her sound reasoning often settled our split decisions on difficult questions" and b) "… offers good backing for her answers and is able to relay that information to her group members. Because of this we have been able to answer many questions that we were in disagreement on correctly." Another way an individual could sway the team appeared related to that individual's ability to question or challenge other's ideas: "He's really good at challenging

the group on a question if he questions it's [sic] accuracy and many times turned group opinion!" Within observed teams, no obvious pattern explained what might be influencing team mates towards or away from the correct persons' answers. High and low performers, quiet and talkative members, as well as trusted and less trusted teammates alike had times when they were unable to convince teammates when they had the right answer. On one quiz, two quiet, lower-performing students in team R, happened to be the only ones with the correct answer, and the team missed the point. However even the best-performing, most talkative and dominant member in that team was not always able to convince the team to choose the correct answer. Likely there are multiple reasons the teams does not choose the correct individual's answer, and these reasons may include both individual characteristics (e.g. performance, dominance, persuasive ability) as well as social dynamics. The group interaction behaviors occurring during those occasions when an individual (or minority) did convince the team may hold the key to identifying critical decision-making behaviors. It is possible that more detailed observations which can link conversations and behaviors to decisions would allow researchers to identify specific group processes that promote or inhibit the correct minority from swaying the group.

Similar to the literature, it appeared that the quantity of members having the same answer may have been one of the strongest forces for choosing an answer in this study. The probability an individual's idea was included in the group solution depended on "the number of individuals who had the same idea *prior* to the group meeting" (Lorge, Fox, Davitz, & Brenner, 1958, p. 364). In 6-member teams working on intellective tasks involving four-term verbal analogy items, three or more correct members were almost certain to influence the group; these groups had a proportion of .90 and higher correct group answers (Laughlin & Adamopoulos, 1980). However when two correct members must persuade up to four incorrect

members to accept the correct answer the group proportion of correct answers drops to .67 and further drops to .28 when only one member was correct. So, groups tended need at least two correct group members for a correct group response. Half of the ideas held in common by two or more group members prior to the group meeting appeared in the group decision, but only 10% of the unique ideas (those from only one person) ultimately appeared in the group solution (Lorge, et al., 1958). Two-thirds of the ideas within the group had been expressed by individuals prior to the group meeting, but one-third of the ideas evolved from within the group (Lorge, et al., 1958).

Differences between tasks. Students exhibited different behaviors when engaged in individual quizzes, group quizzes and clicker questions. Compared to clicker questions, teams spent a greater amount of time in discussion during group quizzes. The teams also more consistently worked together and physically turned to face each other during the discussion during the group quizzes. Comparing group and individual quizzes, students took more time to complete the individual quizzes than the team quizzes. These differences are not surprising since the nature of the task can influence the way that a team moves forward with the task. Hackman and Morris explained how the task "can be designed so that it requires, suggests or provides cues which prompt specific ways of performing the task" (Hackman & Morris, 1975, p. 78). The nature of the task is so important that Michaelsen and Knight devote an entire chapter to explaining how to create effective assignments (Michaelsen, et al., 2004).

Logically, the individual quizzes are the students' first exposure to the items and they needed more time to read the question, think about it, and answer it. For the team quiz, all the questions are the same, so the only new activity is for students to discuss their individual answers and decide on a response. Group decisions could occur quite quickly when team mates had the same answers already. A common procedure used during discussions was for one

member to announce the question number and then each member indicate their answer by merely stating “a”, “b”, “c”, “d” or “e”. If all had the same answer, they would move on to the next item without discussing the substance of agreed-upon questions. Usually, the team only took time to further discuss items if members had different answers.

The group discussions provided an opportunity for team mates to learn from each other and engage in error checking. That process might get short-circuited if members rush through the group quiz with minimal discussion. Robinson (2009) demonstrated that changing the task, even in seemingly minor ways, could have a significant influence on the outcomes. Using a computer-based testing system to change the order of the multiple choice answers on iRATs and gRATs improved solution quality and deliberation time (Robinson, 2009; Robinson, Sweet, & Mayrath, 2008; Robinson & Walker, 2008). Students using different-answer-order iRATs took more time and scored higher when completing the team RAT (using IF-AT forms) than did students who used the same- answer-order iRATs. Presumably, Robinson’s change to the answer arrangement forced students to verbalize their choices rather than simply stating, “I put B,” and hindered students from conducting a quick vote-counting procedure. This change turned questions that might otherwise have received little attention or thought during the discussion into items requiring more care, greater mental work, and “more meaningful consensus-seeking dialogues intended by collaborative learning environments” (Robinson, 2009; Robinson & Walker, 2008, p. 81). Encouraging discussion for every question could be helpful to allow the team to re-analyze the items and perhaps come to a different conclusion. It was outside the scope of this study to determine if learning benefits differed with the two types of interaction. Intuitively, though, the quiz discussions might provide greater learning benefit due to more give-and-take discussion and the opportunity to learn from teammates.

Because different group processes can produce different outcomes, a teacher can best design a task which will produce the intended outcomes when he or she knows how the task design affects team behavior and learning. For example, if the instructor is mostly trying to fine-tune her instruction in real time based on the current "mood" in the class, the clicker questions work great to give the instructor feedback. On the other hand, if the instructor wants students to deeply consider the rationale for a case, a task which encourages discussion, such as the group quiz format, might be the best choice. Ultimately, what would be most helpful for an instructor is to know that an activity with X features will likely generate Y results so that the instructor can choose the activities most appropriate for her goals and purpose.

Considerations

It is not unusual to have unexpected and unanticipated problems arise during research, and this study was no exception. The limitations and problems encountered during this study are explained along with recommendations for future researchers.

Observer effects. Watson and Michaelsen (1988) decided to use observers internal to the group, but Watson et al. (1991) later recommended utilizing external observer reports and descriptive interactive instruments after having experienced "disappointingly inconclusive" results about the relationship between group interaction and decision-making effectiveness when using member self-reports. As recommended, this study utilized external observers as one of several data-collection methods. These external observations were an important component of this study. However, they had limitations, and they may have caused some minor disruptions to teams despite efforts to minimize the impact on those observed.

There were two primary limitations of external observation: 1) the observer's physical presence potentially disrupting normal group activity and 2) the inability of external observers

to have an "insider" perspective. Regarding disruptive physical presence, we presume that the frequent physical presence of the observer near observed teams allowed teams to become habituated to the observer and thus behave "naturally" as they did on days when they were not being observed. Regarding the insider perspective, the observers had limited access to the inner workings of the group, including feelings, tensions, hierarchies, and in-group communication. Thus, the observer(s) had to deduce what teammates felt about each other based on participants' spoken or written responses and body language. Although observers could develop impressions about the group which did not fit the experience from within the team, we attempted to minimize erroneous conclusions through triangulation. When two observers were present, they compared their observations immediately after watching the team. Observer reports were also interpreted along with other data points, such as team members' comments on surveys and peer evaluations. Instances in which external observation impressions conflicted with the message from other data points warned researchers to interpret the findings carefully, but also provided opportunities for closer examination into assumptions about effective group process behaviors. For example, team K did not provide a glowing opinion of their team's functionality, but an observer had felt that this team was the "best teamwork example so far" because team mates appeared to listen to each other and would pick up on an explanation where another left off in the discussion. Also, observers noted a particularly dominant member in team Q who seemed to make two quieter members feel somewhat alienated, but comments from those quiet members were positive and complimentary about the dominant person. These experiences suggest that external observations were not sufficiently sensitive to gauge the group mood or individual feelings. This may have been due, in part, to the physical impossibility of having one or two observers capture important details about the behavior of six team members simultaneously during fast group interactions. With time,

observers noted that some teams' discussions ended in an unexpectedly, abrupt agreement which did not obviously flow from the observed interactions, almost as if members were able to read each other's minds. Presumably there were interactions among team members which were too nuanced or subtle for an outside observer to recognize, and observers may have missed key transitions in the conversation.

I recommend incorporating external observations because they provided valuable insights into the increasing density of group discussions with time, the importance of body language for communication, and the different interaction patterns between quiz discussions and clicker questions. I too think that "even though our results were disappointingly inconclusive, from our own personal observations we strongly believe that the interaction patterns of group members are a key variable with respect to the quality of the decisions that are reached" (Watson, et al., 1991, p. 808). Thus, I recommend that future research supplement external observer reports with additional internal and external observation measures. Video recording could supplement the external observations and help capture fleeting behaviors and conversations that might be missed by a human observer. Researchers could obtain an internal perspective about members' emotional states and thought processes during discussions by conducting interviews immediately after group work, having members answer surveys or keep a journal, or by becoming group participants themselves. Researchers could also verify emerging patterns with group members to ensure accurate interpretations about their feelings when working together, the team decision-making process, and the roles within the group.

Survey method effects. The seating arrangement during survey administration was a potential concern. During the course of this study, several students notified us that the lack of privacy due to students sitting close together while completing the *Teamwork Prediction Questionnaire* could have impacted their honest assessment of nearby peers. Fortunately, the

Teamwork Survey Post-Analysis determined that team mates' proximity had not systematically affected participants' approach to the weekly trust surveys. Future researchers should design the survey response system to ensure that respondents feel confident their responses remain private and secure.

The timing for administering the *Teamwork Prediction Questionnaire* may have contributed to missing data. Students were supposed to finish this weekly survey on their own computers in the time after the individual quiz but before the group discussion. Because the instructor starts the team quiz five minutes after the first team submitted their individual quizzes, some slow persons may not have had time to complete the survey before the team quiz. Future researchers would need to work out such logistical concerns with the instructor to provide a suitable survey-response environment.

Students may not have uniformly interpreted the weekly trust survey. The *Teamwork Prediction Questionnaire* was intentionally designed to measure trust, as defined for this study, in a quickly-completed survey. The literature lists several antecedents to trust and the survey wording was left broad enough to encompass all these various reasons why someone would allow themselves to be vulnerable to another party. The brevity of the survey was less likely to induce survey fatigue, but it sacrificed specificity. The lack of specificity may have caused a few students to be confused about how to respond: "I guess my responses are pretty arbitrary. I dont [sic] think any of use "relies" on anyone else. We are very good at discussing and reaching a decision based on our discussion and everyone contributes." Yet, most of the comments suggest that students did evaluate their teammates on one or more of the antecedents of trust, including competence (i.e. ability to score well on quizzes). Future researchers may wish to ask individuals to rate their teammates on the antecedents of trust from the literature, such as the ten factors (discreetness, integrity, fairness, loyalty, openness, consistency, receptivity,

availability and promise fulfillment) measured by the Conditions of Trust Inventory (Butler, 1991) or the five factors (benevolence, reliability, competence, honesty, openness) described by Tschannen-Moran and Hoy (2000). Such a breakdown of the trust factors could help researchers gain a more nuanced understanding of trust in TBL groups by determining which trustworthiness factors are most important in TBL, which are most closely associated with effective group processes, and what behaviors relate to those trustworthiness factors.

The trust survey instruments used in this study were adapted for use in this study based on existing instruments. The principle component analysis conducted with the Teamwork Process Questionnaire indicated that the items did represent a single construct which I called trust. I confirmed the internal consistency of the measure using Cronbach's Alpha. However, the internal consistency of this measure does not guarantee its test-retest reliability. The Teamwork Prediction Questionnaire was based on one of the questions from the Teamwork Process Questionnaire. These surveys had face validity because they were clear and organized, and they had content validity because the questions fit with characteristics of trust and trust definitions identified in the trust literature. However, these surveys may not have met the requirements for construct validity which ensure that the survey actually measures what it is intended to measure. The confusion indicated by some students as the completed the Teamwork Prediction Questionnaire points to this. Further validation of these instruments would lend greater weight to the findings in this study.

Unique features of veterinary students and TBL teams. To minimize barriers to group cohesiveness, Michaelsen (Michaelsen, 2004a) recommends teachers use "a group formation process that mixes students up in a way that forces the groups to build themselves into teams 'from the ground up'." However, in veterinary medicine, even if teachers create teams appropriately, there is still a possibility that veterinary students exhibit different team

dynamics than observed in TBL studies utilizing undergraduates or non-cohort groups. Veterinary programs tend to generate intimate social environments different than typical undergraduate programs. Veterinary students move through their training as a cohort and they attend the same classes together all day long during the first two or three years. After being "in the trenches" together for extended periods of time, these students get know each other fairly well. Thus, when teams are formed for TBL, the students are not working with strangers, but with people they are familiar to a greater or lesser degree. Rather than building the team from the ground up, the teams could be building teams based on preexisting knowledge of their teammates.

Future study ideas. This observational pilot study revealed several potential patterns in TBL group development and provided a variety of potential avenues for further inquiry. Future research can more closely examine particular patterns, preferably using a controlled experimental design to ensure that causative conclusions are possible.

Replicating seminal TBL studies. During this research, I realized that analyses comparing iRATs and gRATs and making conclusions about the superiority of group performance may be confounded by a practice effect (see Notes, p. 124). Researchers should determine whether a practice effect does influence the member-to-group comparisons such as the group added value scores, synergy scores, and best member to group comparisons. Studies seeking to determine how well team (gRAT) performance compares to individual (iRAT) performance should include appropriate control for the gRATs, such as iRATs taken after group discussion. Ultimately individual learning, not group scores is the important outcome in an educational environment. As such, future TBL research should put more emphasis on how the group process relates to members' depth and breadth of understanding and less on examining how individuals perform compared to the group.

Trust, group development, and group processes (behaviors) in TBL. Assuming TBL teams experienced group development as outlined by sequential group development models, I would have expected trust to drop (during a conflict phase) early in the semester and then rise later in the semester. Instead, various teams had different trust patterns over time. The data from this study generates many questions: How do individual's behaviors affect TBL group processes and learning? How do individuals' behaviors relate to conditions of trust and group performance? Are certain behaviors or breaches of certain conditions of trust more detrimental when expressed by high-performing individuals than low-performing individuals? Are there behaviors associated with the trust antecedents which could be promoting or hindering the group development process? Future studies examining the claim that open discussions improve performance may consider periodically asking members about their willingness to openly share their ideas with team mates while comparing this interview data with observed group interaction behaviors, group development stage and performance over time.

If TBL teams do progress through the developmental stages as described by group development models, it is unclear how long teams remain in each stage, how the stage relates to productivity and whether TBL teams develop faster and achieve more advanced stages of maturity than with other cooperative learning methods. Although research suggests TBL group processes become more effective over time (Birmingham & Michaelsen, 1999; Watson, et al., 1993; Watson, et al., 1991), no known studies have characterized the TBL group development time frame within the framework of established group development models. Assuming that TBL helps most teams quickly become productive (Michaelsen, et al., 2004), teams should be reaching stages three and four before the course ends. After characterizing team development timelines in TBL, researchers may wish to compare group development within TBL to that of other collaborative methods since there is "no direct evidence that groups develop faster

and/or further in TBL classrooms than in classrooms using other kinds of small group learning" (Sweet & Michaelsen, 2007, p. 45). If researchers were to find that TBL was more quickly and consistently shepherding teams through the group development process than other cooperative learning methods, then researchers could examine the TBL structures to identify why.

A first step may be to repeat a modified version of this study. The qualitative data in this study give clues that individual and group behaviors may be influencing decision quality in some cases. Watson et al. (1991) had failed to show a relationship between group interactions and group decision-making effectiveness when utilizing a survey instrument, but their personal observations also suggested that "the interaction patterns of group members are a key variable with respect to the quality of the decisions that are reached." There is important information there and we need to develop appropriate methods to measure it. The modified longitudinal study would measure trust and group development monthly. Participants would be asked to indicate their trust in teammates utilizing an instrument similar to the Conditions of Trust Inventory (Butler, 1991) that breaks trust into specific trustworthiness factors. Team mates would also complete the Group Development Questionnaire (Wheelan & Lisk, 2000) to identify the group development stage. More extensive external observation (videotape) and internal observations (member interviews) should be used to capture all significant interaction events, especially situations when a minority sways the majority or all members answered incorrectly. The study could also measure students' ability to estimate their own ability and participate in group discussions appropriately based on that knowledge. Strategic speaking and accurate self-knowledge might be important characteristics of a meaningful contributor to the group process. Preferably the experiment would be performed in a class utilizing individual and group assessments which are unlikely to have a restricted range. This data could identify the developmental stages occurring in TBL teams and demonstrate the relationship between trust

and developmental stages. The data could also show trustworthiness behaviors and developmental stages which relate to effective group processes and learning. The metacognitive data could show if TBL improved all students' metacognitive ability and if greater metacognitive ability improved team interactions. If researchers can identify which group behaviors are related to better learning in TBL, then instructors could design interventions which promote the most effective behaviors.

Effects of specific changes in course design on performance. It is not uncommon for instructors to modify one or more components of TBL to suit their particular topic or environment. However, there is not much research to indicate how each component affects the outcome and how the components interact together. Thus, instructors are changing the TBL structure without knowing exactly how these alterations will affect the learning in their classroom and may erroneously assume that their results should be equivalent to those in a high-fidelity version of Team Based Learning (Michaelsen & Richards, 2005). Knowing how specific components relate to learning would assist instructional designers in accurately diagnosing learning problems in a TBL course and allow them to strategically modify TBL components to achieve specific outcomes. A good place to start investigation is with the changes to TBL which Michaelsen might consider problematic. The following are a few examples.

Michaelsen has recommended made a number of recommendations that may be worthy of further investigation. When Robert Philpot (2005) asked the TBL community (using the TEAMLEARNING listserv) about potential causes for an odd finding that "some teams actually score lower than the highest member on that team," Michaelsen (2005) gave several explanations including: (a) having a single, short test, (b) giving the tests to newly-formed groups which are not yet teams, (c) not using the IF-AT answer sheets for gRATs and (d) using

an unreliable test. He indicated further that he had not had any team fail to beat its best member since using IF-AT answer sheets (Michaelsen, 2005). In response to another question on the listserv, Michaelsen recommended recording the individual members' attendance and RAT scores in the group folder using an ID number so that the scores were "sort of public within the group" (Michaelsen, 2004b) but also "every team member knows how every other team member is doing" (Michaelsen, 2004a, p. 44; 2004b). The course in this study did not use IF-AT forms nor list the individual's scores inside the team folder. Also, unlike Watson, et al. (1991), who had three 32- to 36-item tests, this course had numerous 5-item quizzes. Future researchers can design simple, randomized controlled experimental trials to determine how certain modifications in TBL, such as the presence or absence of IF-AT answer sheets, the use of individual scores within the team folder, or the number of questions in the RATs, affects group interaction, performance and metacognitive ability.

Appendix A. Instruments

The relevant content for each instrument used in this study is listed below.

Teamwork Process Questionnaire

Please select the response which indicates your level of agreement or disagreement with each statement.

Question	1- Strongly Disagree	2	3	4-Neither agree nor disagree	5	6	7- Strongly Agree
1. We will have confidence in one another on this team.	1	2	3	4	5	6	7
2. I will be able to rely on those I work with in this team.	1	2	3	4	5	6	7
3. There will be a noticeable lack of confidence among those I will work with. *(reversed)*	1	2	3	4	5	6	7
4. Overall, the people in my team will be very trustworthy.	1	2	3	4	5	6	7
5. We will be considerate of one another's feelings in this team.	1	2	3	4	5	6	7
6. The people in my team will be friendly.	1	2	3	4	5	6	7

Teamwork Prediction Questionnaire

For each person on your team, including yourself, please indicate how willing you are to rely on this person for a positive contribution to this group activity.

	Minimally willing to rely on this person (0% to 20%)	**Marginally willing to rely on this person (21% to 40%)**	**Moderately willing to rely on this person (41% to 60%)**	**Generally willing to rely on this person (61% to 80%)**	**Extremely willing to rely on this person (81% to 100%)**
Team member A	1	2	3	4	5
Team member B	1	2	3	4	5
Team member C	1	2	3	4	5
Team member D	1	2	3	4	5
Team member E	1	2	3	4	5
Team member F	1	2	3	4	5
Team member G	1	2	3	4	5

Feel free to comment on your choices:

Teamwork Survey Post-Analysis

1. We are wondering if individuals' responses on the Teamwork survey could have been influenced by the fact that they were sitting next to team members while completing the survey. Did having team members sitting next to you during class influence your responses on the Teamwork Surveys?

☐ No
☐ Yes (Please explain) _____

2. How seriously did you take the Teamwork Survey?

☐ 1. I didn't care what answers I chose
☐ 2
☐ 3
☐ 4. I thought a moderate amount about each choice
☐ 5
☐ 6
☐ 7. I thought hard about each choice I made

Appendix B. Educational Environment Details

The course in this study utilized the following TBL techniques:

Orientating learners to the course format during the first class meetings: The instructor used the first three class days to introduce students to the course format. She explained how she had specifically designed the class, including incorporating Team Based Learning and the Diagnostic Pathfinder, to maximize veterinary student learning. She also guided students through the processes used to complete homework and to submit team choices. Prior to graded assignments, the instructor had students practice the team decision-making processes by using Michaelsen's "Grade-Weight-Setting Exercise" to have them determine what percentage of their grade would come from individual quiz scores, team quiz scores, and peer evaluations.

Separating the class body into permanent, independent, heterogeneous, and appropriately-sized groups: All teams had six or seven teammates, which fits within the recommended team size. The instructor assigned students to teams based on their career and species interests to ensure that teams had diverse resources for group problem solving. Once assigned to their teams, students remained with that team for the entire class.

Mastering the course learning objectives: Because the course assignments were designed to teach and assess the learning objectives in the course, students who passed the course also achieved the course learning objectives by default.

Providing instructional materials for independent learning: The instructor provided students with her clinical pathology notes, which were like a textbook. Students also could purchase a recommended text in physical or electronic form. The latter was linked to the expert diagnostic paths in the Diagnostic Pathfinder homework.

Problem-solving within groups in a classroom environment: Students completed 24 Case Discussion Quizzes and 7 Thursday Team Learning Quizzes. After individual members

completed these activities on their own, the team discussed the items and submitted a team response to the assessment.

Providing opportunities for teams to discuss their rationale with other teams: As students entered their answers to each CDQ, TTQ and clicker question, the selections appeared on screen as frequency graphs, which made disagreements immediately obvious. The instructor encouraged disagreeing teams to explain their reasoning and convince the class why theirs was the correct answer.

Having all teams complete the same assignments: All assessments within the class were the same for all individuals and teams. This included in-lecture clicker questions, CDQs and TTQs. For each question, students were required to simultaneously enter their teams' specific choice into the computer system and the results from the entire class were immediately displayed.

Providing frequent feedback in response to the group discussions: Although the instructor would provide expert feedback in response to questions from the class, the majority of the feedback came from students seeing the distribution of teams' responses after they clicked in. Because the majority response was usually correct, students immediately knew they had chosen the correct or incorrect answer as soon as the frequency graph from all team responses was visible.

Including team performance in grading: The students chose to make team CDQ and TTQ scores 15% of their grade.

Using difficult problems for team assignments: As the semester progressed, the cases became more complex. Mechanisms learned in prior cases were incorporated in the cases students later encountered in both homework and quizzes. For CDQs, students usually had to answer challenging multiple-choice questions about a case. TTQs were more complex because

students had to find the errors intentionally added to the diagnostic path associated with their case.

Requiring peers to evaluate members' contributions to team performance: Students completed two peer evaluations during the semester. The midterm peer evaluations provided practice and opportunity for improvement based on peer feedback. The students chose to make the final peer evaluation scores 10% of their grade.

Requiring individuals to be accountable to the instructor and teammates: The students completed the CDQs and TTQs as individuals prior to the group discussion, and students chose to make their individual CDQ and TTQ scores 15% of their grade. Individuals also received points for completing their homework cases.

There were 4350 minutes theoretically available in the class and students used about 18% of the class time to complete their individual and team quizzes and about 7% of class time when answering clicker questions. The remaining class time was used for reviewing Diagnostic Pathfinder homework prior to quizzes, reviewing quiz answers after quizzes, and lectures. The time for the lecture and review activities was not closely measured; however, rough estimates suggest approximately 51% of class time was spent in lecture and 24% of class time involved homework and quiz review. TBL instructors should have waited until after the RATs to offer "corrective instruction" which addressed concepts the students had been unable to understand on their own because students presumably tend to be "more ready to listen closely to a set of brief, focused statements" after they had struggled with the material (Fink, 2004, p. 10). The post-quiz review does closely fit the "corrective instruction" model described by TBL proponents. The Diagnostic Pathfinder review frequently preceded a quiz. The format, while not involving the timing and format suggested by TBL for "corrective instruction", likely

generated similar urgency and focused attention because the students used the opportunity to clarify their understanding about content before they began the quiz. The lectures had far less urgency, but they too had some corrective instruction value, largely because the intra-lecture clicker questions gave the instructor immediate feedback about the students' misconceptions, and the instructor could immediately adjust the lecture to clarify those issues before moving on. Additionally, if topics seemed particularly confusing to students on their homework or quizzes, the instructor could alter the next scheduled lecture to address those issues.

Detailed Teaching Methods in the Veterinary Clinical Pathology Course

The veterinary pathology course in which I studied Team Based Learning was a comprehensive introduction to clinical pathology for second-year veterinary students and is taught at a large Midwestern University. The instructor had taught Clinical Pathology to veterinary students for 30 years and had begun using Team-Based Learning in 2004. At the time of the study, the instructor used a stable adaptation of the Team Based Learning method which she had fine-tuned over five previous classes using TBL. Although the instructor adhered to most of the principles described in Michaelsens' book, *Team-Based Learning: A Transformative Use of Small Groups in College Teaching* (Michaelsen, et al., 2004), her course design differed to some degree. For comparison, the teaching methods the instructor employed in this class have been described in detail. For ease of reading, this was written as a stand-alone section and, as such, may replicate some information found in the main document.

General Course Characteristics

The Clinical Pathology course was hosted at the veterinary school at one large Midwestern university; and, enrolled in the class were 145 second-year veterinary students from both this institution and another Midwestern university without a veterinary school. The

students attended the class in person and through teleconferencing technologies depending on their location. The class gender distribution was 77% female and 23% male, which was typical for veterinary classes at that time. The instructor divided students into 24 permanent teams with 6 to 7 members each. Each team participated in class activities by using a radio frequency signaling device (nicknamed a "clicker") if they were physically present in class or by using ResponseWare software if they were at a distance. Team selections using these tools were collected into a central computerized system, which enabled the instructor to immediately display students' responses as a graphical summary. Each student had a portable personal computer which they used in class for online quizzes and note-taking. Students accessed copies of lecture PowerPoints, video recordings of previous classes, class schedules, online quizzes and other pertinent information through the courses' WebCT site. The course was taught for a 50 minute session Monday, Tuesday, Wednesday, and Friday, and for a 100-minute session with 10 minute break on Thursday. The class has approximately nine units which lasted from 2 days to 2 weeks depending on the topic. The units included: (1) Erythrocytes; (2) Leukocyte, Monocytes, Lymphocytes, Hematopoietic Neoplasia, Hemostasis; (3) Electrolytes and Acid base; (4) Liver; (5) Urinalysis and Renal; (6) GI; (7) Endocrine; (8) Cytology of Effusions; and (9) Laboratory statistics. The instructor incorporated various interactive teaching techniques into most class periods.

Students received points for completing each graded activity including Diagnostic Pathfinder homework cases, individual quizzes, team quizzes, clicker questions, final peer evaluation scores, and the final examination. These points contributed to students' grades depending on the grade weights. The instructor weighted students' grades so that 50% came from the final exam (1/3 of grade) and the Diagnostic Pathfinder homework (1/6 of grade) and the other 50% was determined by the students. Students set their grade weights for their

individual work, teamwork, and peer evaluations by using the "Grade-Weight-Setting Exercise" described by Michaelsen (Michaelsen; Michaelsen, et al., 2004, pp. 241-248). The instructor automatically dropped the four lowest individual quizzes (CDQ or TTQ) to avoid penalizing students that had to be absent for legitimate reasons.

Team assignment method

The instructor assigned local students to teams on the second class day using a quasi-random method which ensured that each team contained members with diverse professional interests. When forming teams, the instructor projected the following list containing various veterinary career choices or species interests: Small animal practice, Food animal practice, Equine practice, Mixed animal practice, Zoo or wildlife, Exotic companion animal, Research, Academic, Industry position, Government, Other. The instructor announced aloud the topmost category and the students who felt the most connection to this category as their future career choice stood and lined up along the classroom wall. The instructor continued down the list announcing each category one-by-one, and with each category, the students remaining in their seats stood and lined up behind the previous group. Eventually, all students in class were standing in a long line in order of their professional interests around the classroom perimeter. The instructor told the students to sequentially shout out one number from 5 to 24, beginning with the first student in the line. (The first student says "five," and the next says "six." This continues until they get to the person who says "twenty-four," and the next person starts again with "five.") Students with the same number gathered together to form their team and to find their assigned location, where they sat with their team for the semester. Teams 1-4 had been assigned amongst the distance students utilizing a similar method at a prior occasion. All teams except team 5 ended up with six members; Team 5 had seven members.

Teaching strategies used

The instructor melded TBL tools along with teaching techniques that had previously proven successful in her class to actively involve students in learning activities and maximize the learning potential of group work. She used techniques not specifically prescribed by Michaelsen's TBL (dynamic PowerPoint lectures with team-based audience-response questions, Diagnostic Pathfinder homework cases, "food metaphor" demonstrations and student skits) along with several strategies described by Michaelsen (individual and group Case Discussion Quizzes, Thursday Team Learning Quizzes, peer evaluations, appeals, and student-determined grade weights). Each has been described separately below.

Interactive lectures. The instructor used interactive PowerPoint lectures to introduce and explain topics. In the instructor's experience had shown that her students preferred lectures to "preload" the content which students might encounter in their homework cases. The lecture frequently included graphics and animations to illustrate the content. Also, the instructor often interspersed between 1 and 15 "clicker questions" within a lecture. The students, within their teams, discuss the answers to the questions and submit their team response using their team clicker or ResponseWare software. Using the clickers as teams was an adaptation from the instructor's earlier audience-response method which used flashcards presented by each individual. The teams receive one point per correct answer on these "clicker quizzes". The TurningPoint software accepts the responses and displays a graph of responses on the screen. The instructor called on teams selected randomly by drawing balls labeled with team numbers from a canister. She asked team-members to explain their responses and to answer additional questions on related pathophysiology. If there was disagreement about the answer among teams, the instructor did not give away the correct answer but instead encouraged the teams to convince each other using evidence and solid rationales. The in-lecture

clicker questions helped to maintain student attention during lecture, to stimulate discussion and to give the instructor immediate feedback about students' current understanding so that misunderstandings can be quickly corrected.

Food Metaphors and Skits. Occasionally the instructor utilized "Food Metaphors" or skits to illustrate content in the course. Generally food metaphors involved inviting student volunteers to the front to either manipulate or consume food items which represent some aspect about the concept at hand. Skits involved students acting out how various body parts and cells behaved in health and disease. The intention was to provide a visual example which solidifies and simplifies ethereal or complex concepts. For example, one food metaphor involved putting salt, sugar, nail polish remover, sesame seeds and an apple into a sieve to demonstrate what bloodstream components can pass through normal kidney filters into the urine. Student volunteers receive no credit for participating in these activities other than the "fame", learning, and free food. However, the instructor sometimes interspersed clicker questions within and following demonstrations and the teams did receive credit for correctly answering those clicker quizzes.

Diagnostic Pathfinder software. The students utilized the Diagnostic Pathfinder software to complete their homework. This software stepped them through medical cases in which they used historical and laboratory data to generate a "Diagnostic Path", an outline-like rationale for their disease diagnosis. The software included several important features to promote learning: preventing students from skipping over abnormal data, requiring correct medical terminology for abnormal laboratory data, permitting students to group related laboratory data abnormalities in association with underlying causes of disease, and requiring students to utilize or explain each data abnormality in the Diagnostic Path. It was intended to be

a safe place for students to practice developing their methods for working through cases. They completed 72 cases throughout the semester and received one point for each case they submitted. The assigned cases each week related both to the content previously discussed in class and to the in-class assessments that typically followed. New concepts were slowly introduced in each case, with new concepts building upon those previously discussed so that cases became gradually more complex over the semester. When the weekly Diagnostic Pathfinder cases were reviewed in class, students had opportunities to ask the instructor about the case. Often Diagnostic Pathfinder homework cases were reviewed immediately before a quiz on a similar case.

Case Discussion Quizzes. Twenty four unannounced Case Discussion Quizzes (CDQs) were administered throughout the semester, usually after discussing the Diagnostic Pathfinder homework cases which were due that day. Most CDQs consisted of a case scenario with clinical pathology data. The students had to interpret the data and answer five multiple choice questions which probed their ability to recognize the underlying mechanisms and significance of the data. These 5-point assessments had similar topics as to the homework cases, but the quiz contained a different case scenario.

During quiz times, students worked individually on their laptop computers to complete the WebCT quiz released by the instructor. When all team members completed the individual quiz, the team clicked to indicate they had finished. As soon as the first team clicked in, the instructor announced that all teams had five minutes left to complete the quiz. After the individual quiz, the instructor released the team quiz, which exactly replicated the items on the individual quiz. At that point, the team members had the opportunity to discuss their responses and submitted their answer choices as a team. Similar to the individual quiz, the teams used the clicker to indicate they had completed the team quiz. After the first team clicks in, the instructor

announces five minutes for all remaining teams to complete the quiz. When all teams had submitted, the instructor reviewed the quiz question by question and teams clicked in their answer choices. The responses from all teams appeared on the screen as a frequency graph and disagreements were clearly visible as multiple bars on the graph. Disagreeing teams were encouraged to explain their reasoning and to convince the class why theirs was the right answer. Twenty-three of the CDQs used this individual-then-team format, but one was completed only as a team.

Thursday Team Learning Quizzes. The students also engaged in seven Thursday Team Learning quizzes (TTQ) which the instructor had called Team Based Learning activities or TBLs. (In this document, the name of this tool was changed to decrease confusion with Team Based Learning as a teaching system.) For most TTQs, like most CDQs, students received both a case scenario which they had to interpret. However, TTQs also included a completed Diagnostic Path with five intentional errors. TTQs were conducted on Thursdays to allow sufficient time for completion because these were more complex and cognitively demanding than the CDQs. In addition to requiring students to interpret the clinical pathology data for the case, students also had to identify and correct another person's reasoning errors within the Diagnostic Path.

As with the CDQs, the students usually first completed each TTQ as individuals then again as a team. As with CDQs, students had similar opportunities to share team responses with the class and discuss disagreements immediately after completing the team portion. When discussing the TTQ, the instructor would call on a team and ask team members to correct the errors in addition to identifying them. The instructor would make the corrections on the screen as the class progressed through the case.

Appeals. Individuals and teams could appeal missed points on clicker quizzes, CDQs and TTQs by emailing the instructor with a solid rationale explaining why their answers were reasonable. If the instructor accepted the rationale or felt that the student had learned, the instructor granted points back to the team and team members making the appeal.

Final Examination. The cumulative final exam contained 50 multiple choice questions relating to 8 medical cases. Students completed this paper exam as individuals at the end of the semester. No exams are allowed to leave the room to ensure that test questions did not end up in student test banks. The instructor used the same final exam each year.

Peer Evaluations. Students completed online, anonymous peer evaluations halfway during the semester and again at the end of the semester. The mid-term peer evaluation did not contribute to students' grades, but offered teammates the opportunity to provide and receive feedback about teammate's performance. The midterm peer evaluations were also a practice opportunity for the final peer evaluation, which did contribute to student grades. From a given amount of points, each student divided these points among their teammates as they felt appropriate. Each member of 6-person teams has 50 points to divide among teammates, and each member of 7-person teams has 60 points to divide among teammates. The average score received from teammates contributed to the individual's grade. By giving more points to certain members and fewer to others, students could respectively reward or penalize teammates' good or poor performance. Students also assign equal points to all members if they felt all contributed equally. Students were asked to write comments for each team member: "indicate specifically how this person contributes to group success. Your comments will be provided anonymously to the student you are rating." Students received emails containing their midterm and final peer evaluation results respectively. These reports, while not revealing the names of

teammates, did include comments from teammates as well as statistics on the scores this student received from teammates.

Sample Midterm letter:

Dear [student name],

Reported below is the rating and comment data from the VPTH 425 Clinical Pathology midterm peer evaluation. As you know, this rating will not count toward your final grade, but is a practice run for the peer evaluation at the end of the semester.

Your teammates ([# of teammembers who responded] of them) rated your contributions to the success of Team [team #]. You received an average score of **[average of scores given this person by teammates]**. (The overall average score for all students is 10.)

You received the following individual scores:

# scores <9	# scores =9	# scores =10	#scores=11	#scores>11
[# of scores less than 9]	**[# of scores equal to 9]**	**[# of scores equal to 10]**	**[# of scores equal to 11]**	**[# of scores greater than 11]**

Your minimum score was: **[the minimum score this student received from teammates]**

Your maximum score was: **[the minimum score this student received from teammates]**

Comments about how you contribute to group success:

- Comment from Teammate 1
- Comment from Teammate 2
- Comment from Teammate 3
- Comment from Teammate 4
- Comment from Teammate 5
- Comment from Teammate 6 (applicable only for 7-person teams)

Sample final letter:

Dear [student name],

Reported below is the rating and comment data from the VPTH 425 Clinical Pathology final peer evaluation.

Your teammates ([# of teammembers who responded] of them) rated your contributions to the success of Team [team #]. You received an average score of **[average of scores given this person by teammates]**. (The overall average score for all students is 10.)

You received the following individual scores:

# scores <9	# scores =9	# scores =10	#scores=11	#scores>11
[# of scores less than 9]	**[# of scores equal to 9]**	**[# of scores equal to 10]**	**[# of scores equal to 11]**	**[# of scores greater than 11]**

Your minimum score was: **[the minimum score this student received from teammates]**

Your maximum score was: **[the minimum score this student received from teammates]**

Comments about how you contribute to group success:

- Comment from Teammate 1
- Comment from Teammate 2
- Comment from Teammate 3
- Comment from Teammate 4
- Comment from Teammate 5
- Comment from Teammate 6 (applicable only for 7-person teams)

Weights of student work, teamwork and peer evaluations. The instructor assigned the final exam (one-third of grade) and Diagnostic Pathfinder homework (one-sixth of grade) to be worth 50% of students' grades and allowed students to determine the weights for the remaining 50% using the "Grade-Weight-Setting Exercise" (Michaelsen; Michaelsen, et al., 2004, pp. 241-248). Students assigned grade weights for 1) individual work, 2) team work and 3) the final peer evaluation. Individual work included individual CDQ and individual TTQ scores. Teamwork included in-lecture clicker questions, team CDQ scores and team TTQ scores.

The instructor introduced the students to research indicating that (a) most team scores are higher than the highest individual score, (b) individual scores give accountability to

members, and (c) peer scores assure everyone contributes. The instructor asked students to first discuss the weights within their own teams. Then, representatives from each team assembled (distance representatives joined through videoconferencing) and came to consensus on the weights. The representatives returned to the whole class and told the class they chose 25%, 15%, and 10% for individual work, teamwork and peer evaluations respectively. The class agreed and adopted these weights.

Order of events in class

The first and second class days were primarily used to introduce the class format and orient students to the Diagnostic Pathfinder, which students used for their homework throughout the semester. The instructor assigned teams at the end of the second class day. On the third class day, the instructor explained and practiced clicker use, then the students determined the weights for course components which contribute to their grades. For the remainder of the semester, the in-class time alternated between interactive lectures, discussing Diagnostic Pathfinder homework cases, food metaphors, CDQs and TTQs. About 50% of class time was spent in lecture. About 25% of class time was spent reviewing Diagnostic Pathfinder cases and answering student questions before quizzes. About 25% of class time was spent in individual and team work.

Appendix C. Observation Details

Individual performance and trust measures in observed teams

Team	Team Member	Midterm Peer Evaluation	Final peer evaluation	Pre-midterm performance	Post-midterm performance	Semester performance	Trust by team (Semester Average)	Self-trust (Semester Median)	Trust of Team (Semester Median)
Team I	S050	10.00	10.00	3.79	3.69	3.74	4.91	4.00	4.40
	S051	10.00	10.00	4.27	4.23	4.25	4.94	5.00	5.00
	S052	9.75	10.00	3.71	4.18	3.92	4.64	4.00	5.00
	S053	10.25	10.00	4.33	3.85	4.11	4.90	5.00	5.00
	S054	10.00	10.00	4.47	4.18	4.35	4.90	5.00	5.00
	S055	10.00	10.00	3.93	3.62	3.79	4.86	5.00	5.00
Team K	S062	11.40	11.20	4.50	4.31	4.41	4.50	5.00	5.00
	S063	9.00	9.60	3.93	4.17	4.04	4.31	4.00	3.38
	S064	9.40	9.20	4.33	4.27	4.31	4.28	5.00	5.00
	S065	9.40	10.00	4.20	4.54	4.36	4.24	5.00	5.00
	S066	11.00	10.20	3.64	3.80	3.71	4.56	5.00	4.55
	S067	9.80	9.80	3.47	3.38	3.43	4.57	4.00	3.82
Team Q	S098	10.20	9.80	3.73	4.00	3.86	4.52	4.00	3.60
	S099	9.00	9.60	3.73	3.38	3.57	3.49	5.00	5.00
	S100	10.00	10.20	3.60	3.46	3.54	4.20	4.00	5.00
	S101	10.40	10.00	4.40	4.08	4.25	4.72	5.00	3.20
	S102	10.20	10.20	3.67	3.77	3.71	4.49	5.00	4.60
	S103	10.20	10.20	3.33	3.92	3.61	4.17	4.00	5.00
Team R	S104	9.80	10.00	4.20	3.92	4.07	4.74	5.00	5.00
	S105	8.40	9.80	3.15	3.91	3.50	4.46	5.00	5.00
	S106	11.60	10.20	4.40	4.23	4.32	5.00	4.00	4.00
	S107	12.40	10.00	4.00	4.08	4.04	4.76	5.00	5.00
	S108	8.00	10.00	3.64	4.00	3.81	4.64	5.00	5.00
	S109	9.80	10.00	4.13	3.92	4.04	4.72	5.00	5.00
Team V	S128	10.20	10.20	4.27	3.83	4.07	4.74	5.00	5.00
	S129	10.20	10.20	4.47	4.15	4.32	4.77	5.00	5.00
	S130	9.80	9.80	3.87	3.38	3.64	4.93	4.00	4.00
	S131	10.00	10.20	4.53	4.77	4.64	4.68	5.00	5.00
	S132	10.00	9.60	4.33	4.33	4.33	4.75	5.00	5.00
	S133	9.80	10.00	4.80	4.85	4.82	4.82	5.00	5.00

Observation schedule

Each team had 14 or 15 observation data points spread throughout the semester.

Team observation schedule with exam schedule for other courses during the same semester

Time Frame		Monday	Tuesday	Wednesday	Thursday	Friday	Survey
Quarter 1	Week 1			Team 9	Team 18	Team 22	Pre-survey
	Week 2		Team 11	Team 17	Team 22 ‡	Team 9 ‡	Trust survey 1
	Week 3	Team 11	Team 17 ‡	Team 22	Team 9 ‡ (A)	Team 18	Trust survey 2
	Week 4	Team 17 ‡(B)	Team 22	Team 9	Team 18 ‡(C)	Team 11 ‡	Trust survey 3
	Week 5	Team 22 (D)	Team 9 ‡(E)	Team 18 ‡	Team 11	Team 17	Trust survey 4
Quarter 2	Week 6	Team 9 (A)	Team 18	Team 11 ‡	Team 17 ‡	Team 22	Trust survey 5
	Week 7	Team 18 (B)	Team 11 ‡	Team 17	Team 22	Team 9 ○	Trust survey 6
	Week 8	Team 11 (F)	Team 17	Team 22	Team 9 ‡(C)	Team 18 (B)	Trust survey 7
	Week 9	Team 17 (D)	Team 22 (E)	Team 9	Team 18 ‡(A)	Team 11	Trust survey 8
Quarter 3	Mid-Semester Break						
	Week 10	Team 22 ‡	Team 9	Team 18 ‡	Team 11 ○	Team 17	Trust survey 9
	Week 11	Team 9 ‡	Team 18	Team 11	Team 17	Team 22 ○	Trust survey 10
	Week 12	Team 18	Team 11	Team 17 ‡(D)	Team 22 ●(C, A)	Team 9	Trust survey 11
Quarter 4	Week 13	Team 11	Team 17 ‡	Team 22	Team 9 ●	Team 18 ‡	Trust survey 12
	Week 14	Team 17	Team 22	Team 9 ‡	Team 18 ●	Team 11	Trust survey 13
	Week 15	Team 22	Team 9 ‡	Team 18 †	Team 11 ●	Team 17 ‡	Trust survey 14
Final Exams							

Note: The letters within parentheses indicate examination dates for students' concurrent courses: A=Pharmacology, B=Anesthesiology, C=Virology, D=Public Health, E=Surgery, F=Case Studies. The symbols indicate quizzes on those days: ‡ CDQ, individual and team; † CDQ, team only; ● TTQ, individual and team; ○ TTQ, team only.

Schedule and course content difficulty

The difficulty of the material, distractions, and exhaustion could have influenced student performance in this course. During the semester these data were collected, the participants also

attended several other classes which likely drew away students' attention, time and energy from the class studied. As evident in the schedule above, the students were taking many examinations in concurrent courses towards the end of the first quarter and throughout the second quarter which could have resulted in fatigue. The course content is introduced in an increasingly complex, cumulative fashion. Each new case combined new content with review material that involved prior knowledge.

Seating choices and social alliances

I examined the seating choices and social communication interactions because these factors intuitively link to potential social alliances within the teams. Students were not assigned their seats within the team by the instructor, so team members made their own decisions where to sit in relation to other teammates. Students may have selected their seats based on their feelings about certain other members in the team, and they may have chosen to talk with certain members based on their perceptions of those members. In some cases it appeared that persons with whom an individual socialized during class were scored higher by that individual than those in another row, but this was not consistently true. Ultimately, though, the data did not definitively illuminate any pattern between social alliances and peer scores.

Team I, R, Q and V members tended to remain in the same rows the entire semester. In team I, S051, S055 and S053 sat in the upper, back row while S052, S054, and S050 sat in the lower, front row. Although student S050 tended to sit in the same relative position to the other team mates each time, whereas the remaining team mates might shift places within their usual row. In team V, S130, S133, and S131 mostly stayed in the front row and S128, S132 and S129 mostly stayed in the back row. In team Q, S103, S100 and S102 tended to sit in the front and S099, S101, and S098 tended to sit in the back. In team R, S108, S109 and S105 tended to sit in the front and S104, S106, S107 tended to sit in the back except for the first four weeks in the

semester. Initially, student S104 would sit alone one row behind the back row and then this person would join the back row during group activities.

For these teams, the in-class socialization also tended to occur within rows, especially for teams I, R and Q. In team I, the students in the back row would, not infrequently, talk and laugh with each other throughout class and they appeared to have an amicable relationship with each other. Occasionally S054 and S052 would talk to each other during class in a subdued manner. Student S050 was frequently quiet both during class and during discussions. In team Q, the team members frequently chatted and joked loudly with each other and they often spent much time in extended discussions explaining the answers to each other, even after clicking in the response. S098 and S101 would frequently chat in the back, and S100 would sometimes chat with S102 in the front. In Team R, S106 tended to talk with the others in the back row, and in the front S109 and S108 would occasionally talk together. While it is possible that these social alliances influenced peer scores, the picture is mixed. In some cases it appeared that persons who an individual socialized with during class were scored higher by that individual than those in another row, but this was not consistently true.

In team K, the members, for the most part, did not maintain self-assigned seating common in the other observed teams. Except for S062, who mostly sat in the front, and S063, who mostly sat in the back, the team members frequently sat in different seats relative to each other. During class time, team K members would occasionally chat with persons happened to be sitting next to them, whether they were team mates or not. Team members S064, S063, and S067 tended to be quietest members.

Notes

I initially planned to use *Synergy* values and *Group Added Values* in this study in order to replicate the work done by Watson, et al. (1991) and to determine how efficiently groups utilized their resources over time as a function of trust. Presumably, increasing positive synergy and group added values suggested that groups outperformed individuals due to a synergistic benefit enjoyed by groups and that groups become increasingly proficient at generating high quality decisions as members became more familiar with each other over time (Michaelsen, et al., 1989; Watson, et al., 1991). By this logic it is more advantageous for individuals to work within a team because the team can achieve more points than any particular individual by themselves.

Because several problems quickly became apparent with these measures, these analyses were abandoned in this study. To some, the lack of these measures may be a glaring absence, so I report here my intended use, findings and conclusions about these measures. The *group added value* and *synergy* scores were both performance measures which were intended to indicate the contributions of the team over the best individual. *Group Added Value (GAV)* = (Group score - high member score)/high member score. *Synergy Ratio* = (Group score – high member score) / (maximum possible score - high member score). Regarding the synergy ratio, Watson et al. describe the denominator as indicating "how much of a gain could have been added beyond the best member's input" and the numerator as indicating "how much of a gain actually was added by reaching a group consensus" (1991, p. 805).

A value of "0" indicated that the group and individual had equal scores. Negative values indicated that the high individual outperformed the group, "indicating process loss" (Watson, et al., 1991, p. 805). Positive values occurred when the group outperformed the high individual in the group. A synergy ratio of +1 meant that the group achieved a perfect score despite all team

members missing at least one question. However, the synergy formula generated errors whenever the high member achieved the maximum score due to a 0 in the denominator. This was not a problem for Watson et al. (Watson, et al., 1991) because no individual or team achieved a perfect score in their study; however, our study not uncommonly had high members achieving the maximum possible points.

Rather than attempt to adjust the formula to account for this problem, these measures were eliminated entirely. For the purpose of determining how much teamwork provides a learning advantage over individual work, these measures are flawed when used with standard Team Based Learning classroom data. First, this comparison focuses too much on how well the group score and tells us less about how well the members, especially the worst performing members, are learning in the class. Ultimately individual learning, not group scores is the important outcome in an academic environment. Comparing group to best individual distracts from the truly important question of whether working in the group had helped each member in that group learn better than they would have individually. As such, future TBL research should put more emphasis on how the group process relates to members' depth and breadth of understanding and less on examining how individuals perform compared to the group.

Second, using the measures with data from a standard TBL classroom provides inadequate experimental controls. Studies have shown that groups can have superior performance to individuals due to unique, beneficial features that individuals lack, such as an error-correction function and resource pooling (Hill, 1982). But, one cannot accurately draw conclusions about group superiority merely by comparing gRATs to iRATs from standard TBL classrooms because such analyses fail to account for the inherent practice effect which may unevenly benefit groups over individuals. In TBL individuals take the test, the group discusses, then the group re-takes the assessment as a group. Thus, in addition to any group-related

benefits, the group quiz scores also reflect the additional practice (learning opportunities) that occurred during the individual quiz and the discussion. This confounds the analysis. To illuminate the benefits of teams, a better, but still less-than-ideal, comparison would include the final scores individuals and groups who both had opportunities to discuss:

1. Pre-iRAT → group discussion → gRAT
2. Pre-iRAT → group discussion → iRAT
3. Pre-iRAT → self-study → iRAT

The post-treatment iRATs and gRATs could then be compared on more equal footing. This would better illuminate the benefits of teams with less interference from a practice effect.

References

Abdelkhalek, N., Hussein, A., Gibbs, T., & Hamdy, H. (2010). Using team-based learning to prepare medical students for future problem-based learning. *Medical Teacher, 32*(2), 123-129. doi: 10.3109/01421590903548539

Amaria, R. P., Biran, L. A., & Leith, G. O. M. (1969). Individual versus co-operative learning 1. Influence of intelligence and sex. *Educational Research, 11*(2), 95-103. doi: 10.1080/0013188690110202

Arrow, H., Poole, M. S., Henry, K. B., Wheelan, S., & Moreland, R. (2004). Time, Change, and Development. *Small Group Research, 35*(1), 73-105. doi: 10.1177/1046496403259757

Arum, R., & Roksa, J. (2011). *Academically Adrift: Limited Learning on College Campuses*. Chicago: University of Chicago Press.

Ashman, A. F., & Gillies, R. M. (1997). Children's cooperative behavior and interactions in trained and untrained work groups in regular classrooms. *Journal of School Psychology, 35*(3), 261-279.

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).

Bhattacharya, R., Devinney, T. M., & Pillutla, M. M. (1998). A Formal Model of Trust Based on Outcomes. *The Academy of Management Review, 23*(3), 459-472.

Bigley, G. A., & Pearce, J. L. (1998). Straining for Shared Meaning in Organization Science: Problems of Trust and Distrust. *The Academy of Management Review, 23*(3), 405-421.

Birmingham, C., & McCord, M. (2004). Group Process Research: Implications for Using Learning Groups. In L. K. Michaelsen, A. B. Knight & L. D. Fink (Eds.), *Team-based learning: A*

Transformative Use of Small Groups in College Teaching (pp. 73-93). Sterling, VA: Stylus Publishing, LLC.

Birmingham, C., & Michaelsen, L. K. (1999). *Conflict resolution in decision making teams: a longitudinal study.* Paper presented at the Midwest Academy of Management, Chicago, IL.

Bray, R. M., Kerr, N. L., & Atkin, R. S. (1978). Effects of group size, problem difficulty, and sex on group performance and member reactions. *Journal of Personality and Social Psychology, 36*(11), 1224-1240. doi: 10.1037/0022-3514.36.11.1224

Butler, J. K. (1991). Toward Understanding and Measuring Conditions of Trust: Evolution of a Conditions of Trust Inventory. *Journal of Management, 17*(3), 643-663.

Butler, J. K. (1992). Scales and Items of the Conditions of Trust Inventory.

Colbeck, C. L., Campbell, S. E., & Bjorklund, S. A. (2000). Grouping in the Dark: What College Students Learn from Group Projects. *The Journal of Higher Education, 71*(1), 60-83.

Conway, S. E., Johnson, J. L., & Ripley, T. L. (2010). Integration of Team-Based Learning Strategies Into a Cardiovascular Module. *American Journal of Pharmaceutical Education, 74*(2), 7. doi: 10.5688/aj740235

Corbitt, G., Gardiner, L. R., & Wright, L. K. (2004). *A Comparison of Team Developmental Stages, Trust and Performance for Virtual versus Face-to-Face Teams.* Paper presented at the 37th Annual Hawaii International Conference on System Sciences, Waikoloa, HI.

Correia, A.-P., Karpova, E., & Baran, E. (2008). *Questionnaire I: Processes of Teamwork* Analysis of Trust Building in Virtual Teams. Manuscript in preparation. unpublished. Ames, IA.

Covey, S. M. R., & Merrill, R. R. (2006). *The speed of trust : the one thing that changes everything*. New York: Free Press.

Dunaway, G. A. (2005). Adaption of team learning to an introductory graduate pharmacology course. [Article]. *Teaching and Learning in Medicine, 17*(1), 56-62.

Feichtner, S. B., & Davis, E. A. (1984). Why Some Groups Fail: A Survey of Students' Experiences with Learning Groups. *Journal of Management Education, 9*(4), 58-73. doi: 10.1177/105256298400900409

Feingold, C. E., Cobb, M. D., Givens, R. H., Arnold, J., Joslin, S., & Keller, J. L. (2008). Student perceptions of team learning in nursing education. *Journal of Nursing Education, 47*(5), 214-222.

Fink, L. D. (2004). Beyond Small Groups: Harnessing the Extraordinary Power of Learning Teams. In L. K. Michaelsen, A. B. Knight & L. D. Fink (Eds.), *Team-based learning: A Transformative Use of Small Groups in College Teaching* (pp. 3-26). Sterling, VA: Stylus Publishing, LLC.

Foldy, E. G., & Buckley, T. R. (2010). Re-Creating Street-Level Practice: The Role of Routines, Work Groups, and Team Learning. *Journal of Public Administration Research and Theory, 20*(1), 23-52. doi: 10.1093/jopart/mun034

Fox, D. J., & Lorge, I. (1962). The relative quality of decisions written by individuals and by groups as the available time for problem solving is increased. *The Journal of Social Psychology, 57*(1), 227-242.

Gagné, R. M., Briggs, L. J., & Wager, W. W. (1992). *Principles of Instructional Design* (4th ed.). Belmont, CA: Wadsworth/Thomson Learning.

Gulati, R. (1995). Does familiarity breed trust? The implications of repeated ties for contractual choice in alliances. *Academy of Management Journal, 38*(1), 85-112.

Gulati, R., & Sytch, M. (2008). Does familiarity breed trust? Revisiting the antecedents of trust. *Managerial and Decision Economics, 29*(2-3), 165-190.

Hackman, J. R. (1983). Group Influences on Individuals. In M. D. Dunnette (Ed.), *Handbook of Industrial and Organizational Psychology* (pp. 1455-1525). New York: Wiley.

Hackman, J. R., & Morris, C. G. (1975). Group Tasks, Group Interaction Process, and Group Performance Effectiveness: A Review and Proposed Integration. In B. Leonard (Ed.), *Advances in Experimental Social Psychology* (Vol. 8, pp. 45-99). New York: Academic Press.

Hackman, J. R., & Vidmar, N. (1970). Effects of Size and Task Type on Group Performance and Member Reactions. *Sociometry, 33*(1), 37-54.

Hill, G. W. (1982). Group Versus Individual Performance: Are N + 1 Heads Better Than One? *Psychological Bulletin, 91*(3), 517-539.

Jarvenpaa, S. L., Shaw, T. R., & Staples, D. S. (2004). Toward Contextualized Theories of Trust: The Role of Trust in Global Virtual Teams. *Information Systems Research, 15*(3), 250–267.

Johnson, D. W., Johnson, R. T., & Scott, L. (1978). The effects of cooperative and individualized instruction on student attitudes and achievement. [Article]. *Journal of Social Psychology, 104*(2), 207.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative Learning Returns To College What Evidence Is There That It Works? *Change: The Magazine of Higher Learning, 30*(4), 26 - 35. doi: 10.1080/00091389809602629

Jones, G. R., & George, J. M. (1998). The Experience and Evolution of Trust: Implications for Cooperation and Teamwork. *The Academy of Management Review, 23*(3), 531-546.

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.

Lane, D. R. (2008). Teaching skills for facilitating team-based learning. *New Directions for Teaching and Learning, 2008*(116), 55-68. Retrieved from http://dx.doi.org/10.1002/tl.333 doi:10.1002/tl.333

Laughlin, P. R. (1973). Focusing strategy in concept attainment as a function of instructions and task complexity. *Journal of Experimental Psychology, 98*(2), 320-327.

Laughlin, P. R., & Adamopoulos, J. (1980). Social combination processes and individual learning for six-person cooperative groups on an intellective task *Journal of Personality and Social Psychology, 38*(6), 941-947. doi: 10.1037/0022-3514.38.6.941

Laughlin, P. R., & Bitz, D. S. (1975). Individual versus dyadic performance on a disjunctive task as a function of initial ability level. *Journal of Personality and Social Psychology, 31*(3), 487-496. doi: 10.1037/h0076475

Lisk, A. R. (2003, April 5). *Using Teams To Enhance Learning In Management Education.* Paper presented at the Annual Conference of the Midwest Academy of Management, St. Louis, MO.

Lorge, I., Fox, D., Davitz, J., & Brenner, M. (1958). A survey of studies contrasting the quality of group performance and individual performance, 1920-1957. *Psychological Bulletin, 55*(6), 337-372. doi: 10.1037/h0042344

Maccoby, M. (2003). To build trust, ethics are not enough. *Research-Technology Management, 46*(5), 59-60.

Mayer, R., Davis, J. H., & Schoorman, D. F. (1995). An integrative model of organizational trust. *The Academy of Management Review, 20*(3), 709-734.

Meeuwsen, H. (2003). Using team-based learning strategies in undergraduate motor learning and control courses: A continuation *Journal of Sport & Exercise Psychology, 25 (Suppl.)*, S99.

Meeuwsen, H. J., & Pederson, R. (2006). Group Cohesion in Team-Based Learning. *MountainRise Journal, 3*(1), 21. Retrieved from http://mountainrise.wcu.edu/index.php/MtnRise/article/view/88

Merriam, S. B. (2002). *Qualitative research in practice: examples for discussion and analysis* (1st ed.). San Francisco: Jossey-Bass.

Michaelsen, L. K. Getting Started with Team-based Learning Retrieved June 28, 2010, from http://teambasedlearning.apsc.ubc.ca/d/GettingStartedwithTBL.pdf

Michaelsen, L. K. (2004a). Getting Started with Team-Based Learning. In L. K. Michaelsen, A. B. Knight & L. D. Fink (Eds.), *Team-based learning: A Transformative Use of Small Groups in College Teaching* (pp. 27-50). Sterling, VA: Stylus Publishing, LLC.

Michaelsen, L. K. (2004b, September 9). Re: Should teams know each members individual scores?. [Electronic mailing list message]. Retrieved from http://list.olt.ubc.ca/cgi-bin/wa?A2=ind0409&L=TEAMLEARNING-L&P=R819&Y=jim.sibley%40ubc.ca

Michaelsen, L. K. (2005, Nov 24). Re: TBL statistical analysis. [Electronic mailing list message]. Retrieved from http://list.olt.ubc.ca/cgi-bin/wa?A2=ind0511&L=TEAMLEARNING-L&P=R1388&Y=jim.sibley%40ubc.ca

Michaelsen, L. K., & Fink, L. D. (2008). Preface. *New Directions for Teaching and Learning, 2008*(116), 1-5. Retrieved from http://dx.doi.org/10.1002/tl.329 doi:10.1002/tl.329

Michaelsen, L. K., Fink, L. D., & Knight, A. (n.d.). Designing Effective Group Activities: Lessons for Classroom Teaching and Faculty Development Retrieved July 27, 2010, from http://teambasedlearning.apsc.ubc.ca/d/activitydesign.pdf

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 Publishing, LLC.

Michaelsen, L. K., & Richards, B. (2005). Drawing conclusions from the team-learning literature in health-sciences education: A commentary. [Commentary]. *Teaching and Learning in Medicine, 17*(1), 85-88.

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

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.

Naik, A. D., Teal, C. R., Rodriguez, E., & Haidet, P. (2011). Knowing the ABCs: A comparative effectiveness study of two methods of diabetes education. *Patient Education and Counseling, 85*(3), 383-389. doi: 10.1016/j.pec.2011.01.010

Nannestad, P. (2008). What have we learned about generalized trust, if anything? *Annual Review of Political Science, 11*, 413-436. doi: 10.1146/annurev.polisci.11.060606.135412

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

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3 ed.). Thousand Oaks, Calif.: Sage Publications.

Philpot, R. J. (2005, Nov 23). Re: TBL statistical analysis. [Electronic mailing list message]. Retrieved from http://list.olt.ubc.ca/cgi-bin/wa?A2=ind0511&L=TEAMLEARNING-L&P=R142&Y=jim.sibley%40ubc.ca

Phye, G. (1992). Strategic transfer: A tool for academic problem solving. *Educational Psychology Review, 4*(4), 393-421. doi: 10.1007/bf01332145

Phye, G. D. (1997a). Classroom Assessment: A Multidimensional Perspective. In G. D. Phye (Ed.), *Handbook of Classroom Assessment: Learning, Adjustment, and Achievement* (pp. 33-51). San Diego: Academic Press.

Phye, G. D. (1997b). Inductive Reasoning and Problem Solving: The Early Grades. In G. D. Phye (Ed.), *Handbook of Academic Learning: Construction of Knowledge* (pp. 451-471). San Diego: Academic Press.

Phye, G. D. (1997c). Learning and Remembering: The Basis for Personal Knowledge Construction. In G. D. Phye (Ed.), *Handbook of Academic Learning: Construction of knowledge* (pp. 47-64). San Diego: Academic Press.

Phye, G. D., & Sanders, C. E. (1992). Accessing strategic knowledge: Individual differences in procedural and strategy transfer. *Contemporary Educational Psychology, 17*(3), 211-223. doi: Doi: 10.1016/0361-476x(92)90061-3

Pileggi, R., & O'Neill, P. N. (2008). Team-Based Learning Using an Audience Response System: An Innovative Method of Teaching Diagnosis to Undergraduate Dental Students. *Journal of Dental Education, 72*(10), 1182-1188.

Robinson, D. H. (2009). *Using alternate forms to increase deliberation and improve team RAT performance*. Paper presented at the Annual meeting of the Team-based Learning Conference, Austin, TX.

Robinson, D. H., Sweet, M., & Mayrath, M. (2008). A computer-based, team-based testing system. In D. H. Robinson & G. Schraw (Eds.), Recent Innovations in Educational Technology that Facilitate Student Learning In D. H. Robinson, J. M. Royer & G. Schraw (Series Eds.), Current Perspectives on Cognition, Learning and Instruction (pp. 277-290). Charlotte, NC: Information Age Publishing.

Robinson, D. H., & Walker, J. D. (2008). Technological alternatives to paper-based components of team-based learning. *New Directions for Teaching and Learning, 2008*(116), 79-85. Retrieved from http://dx.doi.org/10.1002/tl.335 doi:10.1002/tl.335

Sheppard, B. H., & Sherman, D. M. (1998). The Grammars of Trust: A Model and General Implications. *The Academy of Management Review, 23*(3), 422-437.

Skon, L., Johnson, D. W., & Johnson, R. T. (1981). Cooperative peer interaction versus individual competition and individualistic efforts: Effects on the acquisition of cognitive reasoning strategies. *Journal of Educational Psychology, 73*(1), 83-92. doi: Doi: 10.1037/0022-0663.73.1.83

Smith, P. L., & Ragan, T. J. (2005). *Instructional design* (3rd ed.). Hoboken, N.J.: J. Wiley & Sons.

Springer, L., Stanne, M. E., & Donovan, S. (1997). *Effects of Small-Group Learning on Undergraduates in Science, Mathematics, Engineering, and Technology: A Meta-Analysis.* . Paper presented at the 22nd Annual Meeting of the Association for the Study of Higher Education, Albuquerque, NM. http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED415814

Stahl, R. J. (1994). The Essential Elements of Cooperative Learning in the Classroom (pp. 8). Bloomington, IN: ERIC Clearinghouse for Social Studies/Social Science Education

Sweet, M. (2008). *Individual and Peer Calibration in Team-Based Testing.* Doctor of Philosophy, University of Texas, Austin, TX.

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

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

Tschannen-Moran, M., & Hoy, W. K. (2000). A multidisciplinary analysis of the nature, meaning, and measurement of trust. *Review of educational research, 70*(4), 547-593.

Tuckman, B. W. (1965). Developmental Sequence in Small Groups. *Psychological Bulletin, 63*(6), 384-399.

Tuckman, B. W., & Jensen, M. A. (1977). Stages of small-group development revisited. *Group & Organization Studies, 2*(4), 419-427.

Van De Ven, A. H., & Delbecq, A. L. (1974). The Effectiveness of Nominal, Delphi, and Interacting Group Decision Making Processes. *The Academy of Management Journal, 17*(4), 605-621.

van Woerkom, M., & Croon, M. (2009). The relationships between team learning activities and team performance. *Personnel Review, 38*(5), 560-577. doi: 10.1108/00483480910978054

Watson, W. E., Cooper, D., Torres, M. A. J. L. N., & Boyd, N. G. (2008). Team processes, team conflict, team outcomes, and gender: An examination of US and Mexican learning teams. *International Journal of Intercultural Relations, 32*(6), 524-537. doi: 10.1016/j.ijintrel.2008.01.002

Watson, W. E., Johnson, L., Kumar, K., & Critelli, J. (1998). Process gain and process loss: Comparing interpersonal processes and performance of culturally diverse and non-

diverse teams across time. *International Journal of Intercultural Relations, 22*(4), 409-430.

Watson, W. E., Johnson, L., & Merritt, D. (1998). Team orientation, self-orientation, and diversity in task groups: Their connection to team performance over time. *Group & Organization Management, 23*(2), 161-188.

Watson, W. E., Kumar, K., & Michaelsen, L. K. (1993). Cultural Diversity's Impact on Interaction Process and Performance: Comparing Homogeneous and Diverse Task Groups. *Academy of Management Journal, 36*(3), 590-602.

Watson, W. E., & Michaelsen, L. K. (1988). Group interaction behaviors that affect group performance on an intellective task. *Group & Organization Studies, 13*(4), 495-516.

Watson, W. E., Michaelsen, L. K., & Sharp, W. (1991). Member competence, group interaction, and group decision making: A longitudinal study. *Journal of Applied Psychology, 76*(6), 803-809.

Wheelan, S. A. (2005). *Group Processes: A Developmental Perspective* (2nd ed.). Boston: Allyn and Bacon.

Wheelan, S. A. (2009). Group Size, Group Development, and Group Productivity. *Small Group Research, 40*(2), 247-262. doi: 10.1177/1046496408328703

Wheelan, S. A., Davidson, B., & Tilin, F. (2003). Group Development Across Time: Reality or Illusion? *Small Group Research, 34*(2), 223-245. doi: 10.1177/1046496403251608

Wheelan, S. A., & Krasick, C. L. (1993). The emergence, transmission, and acceptance of themes in a temporary system of interacting groups. *Group and Organization Management, 18*(2), 237-260.

Wheelan, S. A., & Lisk, A. R. (2000). Cohort Group Effectiveness and the Educational Achievement of Adult Undergraduate Students. *Small Group Research, 31*(6), 724-738. doi: 10.1177/104649640003100605

Wheelan, S. A., & Williams, T. (2003). Mapping Dynamic Interaction Patterns in Work Groups. *Small Group Research, 34*(4), 443-467. doi: 10.1177/1046496403254043

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 & Learning in Medicine, 23*(1), 28-30. doi: 10.1080/10401334.2011.536756

Woolley, A. W., Gerbasi, M. E., Chabris, C. F., Kosslyn, S. M., & Hackman, J. R. (2008). Bringing in the experts: how team composition and collaborative planning jointly shape analytic effectiveness. *Small Group Research, 39*(3), 352. doi: 10.1177/1046496408317792

Zgheib, N. K., Simaan, J. A., & Sabra, R. (2010). Using team-based learning to teach pharmacology to second year medical students improves student performance. *Medical Teacher, 32*(2), 130-135. doi: 10.3109/01421590903548521