Team-based Learning, Team Composition, and Team Success: What Really Matters?

Molly Espey
John E. Walker Department of Economics
Clemson University
Clemson, SC 29634
864-656-6401
mespey@clemson.edu

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2012 AAEA Annual Meeting, Seattle, Washington, August 12-14, 2012

Introduction:

The National Association of Colleges and Employers annual survey of employers consistently finds "teamwork skills" and the ability to work with others among the top five qualities employers most desire in employees. Good teamwork requires effective communication, regular interaction, mutual respect and trust. Successful use of teams in the classroom can improve student motivation (Denton 1994; Dommeyer 1986), and increase subject matter knowledge and critical thinking skills (Nichols 2002; Espey 2012), improve communication skills (Meyer 1994).

Unfortunately, students' most common experience working with others in academic settings is either in group activities with groups changing from activity to activity or in groups assigned for one-time projects, with the work primarily conducted outside class time. Such group activities are vulnerable to the pitfalls of free riding and social loafing and may also encounter problems associated with personality differences and inexperience with conflict resolution. Team-based learning (TBL) is an instructional strategy designed to minimize the influence of many of these potential problems. Within the context of TBL, groups are transformed into teams, ideally comprised of a small number of people with complementary skills working cooperatively to achieve a common goal and hold themselves mutually accountable (Michaelsen et al 2002). What constitutes "complementary skills", however, is not always apparent. Identifying characteristics of the structure of teams that contribute to team success could enhance student success in team-based activities in courses. Increased success arguably correlates positively with improved attitudes and ability to work with others in subsequent activities, be they in the classroom or in a work environment.

This study reviews the literature on team composition and team effectiveness, provides an overview of TBL, then presents the analysis of six years' worth of TBL teams in introductory microeconomic theory classes to determine what characteristics of the composition of teams contributes to team cohesiveness and team success.

Background:

The literature related to team composition and team effectiveness tends to fall into two general categories: teamwork in organizations and teamwork in specific course activities. Within the organizational behavior literature, team effectiveness is commonly modeled using an input-process-output framework (Cummings 1981; Hackman 1987; Salas et al 1992) where various input factors such as features of the group, its task, and the work environment influence the group interaction process, ultimately influencing the group output (Hackman 1987). In a work setting, individuals are typically assigned to teams based on availability and competencies relevant to the task at hand. Occasionally, members may volunteer to participate and it may also be possible for individuals to opt out of certain team assignments. In contrast, in the classroom, all students must participate and will generally face the same task and work environment, at least within a given class. Nonetheless, a variety of individual level, team level, and environmental factors may influence team effectiveness (Hackman 1987; McGrath 1964; Yeatts and Hyten 1998).

Surface level variables are overt demographic characteristics such as age, race, education level, and gender. Demographic heterogeneity may lead to differences in team performance

(Bunderson and Sutcliffe 2002; Pelled et all 1999), but meta-analysis has not found a consistent, statistically significant relationship between such heterogeneity and team performance (Webber and Donahue 2001). Deep level variables that may influence team performance include personality factors, values, abilities, and attitudes toward the task at hand and have generally found to have a more significant impact on team success over time (Bell 2007). However, these variables are typically much more difficult to measure than demographic variables and previous empirical results have been inconsistent in terms of which variables matter most (Bell 2007), with only general mental ability (GMA) having been found in two meta-analyses as a strong predictor (Devine and Phillips 2001; Stewart 2006).

The influence of team size has also been investigated with mixed results. Team size is posited to influence the distribution of participation and the nature of group interactions (Dawe 1934; Miller 1951), with smaller teams providing more opportunity to interact, yet other research has found that larger groups perform at higher levels (Campion et al 1993). Cossé et al (1999) found team size to be positively correlated with team performance in a marketing class, but Deeter-Schmelz et al (2011) did not find size to be significant in their study of marketing classes. Class size does not seem to have been studied in terms of its influence on individual team performance in the classroom.

Team-Based Learning:

Team-based learning (TBL) attempts to foster effective group interaction by keeping students in the same group throughout the semester and utilizing collaborative activities daily in class. In terms of team effectiveness, the TBL literature tends to focus on team size and regular interaction as keys to team success, emphasizing team sizes of 5 to 7 students and daily in-class interaction (Michaelsen et al 2002; Michaelsen and Sweet 2008). In such a context, "teams" become distinct from and more effective than "groups." Over time, as students begin to trust each other and develop a commitment to the group, the group becomes a team (Michaelsen et al, 2002).

Sweet and Michaelsen (2011) further describe four pieces of "the practical framework of TBL" as proper teams, the readiness assessment process, 4-S application exercises, and student peer evaluations. In addition to being permanent teams of 5 to 7 students, proper teams are strategically formed by the instructor to provide a balance of resources across teams; this may be done by surveying students on the first day of class regarding prior experience relevant to the class, previous coursework, or any other features that may contribute to success in the specific course. The readiness assessment process involves beginning-of-unit readiness assessment tests (RATs) over assigned reading which are taken individually, then again by teams. This holds individuals accountable for doing the assigned readings and teams accountable for working together to ensure students understanding the basic concepts well enough to begin applying that knowledge. 4-S application activities are significant, involve specific choices, include the same problem for all students, and require simultaneous reporting. Significant problems engage students in concrete examples so they understand the usefulness of the course concepts. Specific choices require teams to take a position, sometimes also requiring them to support that position with a short rationale for their choice. Forcing all students to confront the same problem enables them to better engage with each other across teams, while simultaneous reporting precludes teams from simply agreeing with the majority of others, forcing them to decide before knowing what other groups will say.

In the typical TBL class, student grades are based on both individual work and teamwork. Graded teamwork includes both the team RATs and application exercises, although the number of graded versus ungraded application activities will vary, as will the weight given to each component of the grade. The final critical piece of TBL is student peer evaluation. Again, there is variation in how peer evaluation is implemented across instructors, but it will generally involve both quantitative and qualitative evaluation, be anonymous, and include a mid-semester formative assessment, as well as end-of-semester summative assessment. This peer evaluation will factor into student grades as a third component or used to weight the team portion of the grade.

TBL is designed to address many of the pitfalls of group work. Free riding and social loafing can be minimized through effective us of mid-semester formative peer evaluations and end-of-semester summative peer evaluations. Personality or cultural differences and inexperience with conflict resolution can be mitigated over time, with team performance generally increasing with hours of interaction for at least 30 hours (Watson et al 1993), close to the length of time students have to work together during a regular course. Nonetheless, differences in academic capabilities, attitudes, and effort, and demographic heterogeneity may still play an important role in influencing team cohesiveness and team performance.

Data:

Team performances of 89 teams, comprised of a total of 537 students in twelve sections of introductory microeconomic theory taught between 2006 and 2011, were analyzed to determine what observable characteristics of teams appear to influence team success. Surface-level variables available for each student include gender, class level, major, and where the student is from. Deep-level variables include grade point average, used to measure students' level of effort and/or value of education, and team cohesiveness, measured by variance in end-of-semester peer evaluations for each team. Control or environmental variables include team size and class size. This data is summarized in Table 1.

About 43 percent of the students were female, about a quarter from out of state, and nearly two-thirds were freshmen or sophomores, as might be expected for an introductory level course. Since fewer than 10 percent of the students are actually applied economics majors, the "major" variable measured the percentage of students on each team for whom the course is required by their major. Other students enroll in the course to satisfy the social science general education requirement of the university. While this variable ranged from zero to 100, the average was quite high at 81 percent. Several different measures of grades were used in the analysis: the highest individual's grade point average (GPA) for each team to determine how much influence one person might have on the outcome, the average GPA of all the individuals on the team, and the lowest individual GPA on the team.

As TBL guidelines recommend that teams range from five to seven in size, all but one team in this study fall in that range, with an overall average size of six. In one small class, a student dropped to far into the semester to make team size adjustments, resulting in a team of four. Most of the classes were between 19 and 60 students in size, with the exception of one class of 103. Thus, while the average class size is 46.6, omitting the class of 103 results in an average of 39.5.

Team grades include beginning-of-unit readiness assessment tests (RATs) and in-class team activities. Teams also worked on numerous ungraded activities daily. Over the course of

the semester, teams took five RATs covering basic concepts related to the readings for each unit. These tests were taken first as individuals, then as teams. Each team also completed 12 to 13 graded in-class activities over the course of the semester. The overall team grade was a weighted average of these two components, 25-30% weight given to the RATs and 70-75% weight given to the activities. Individuals' grades were a weighted average of individual activities (homework and tests) and the team grade, weighted by the peer evaluations.

Peer evaluations completed at the end of the semester require students to rate their teammates in terms of their contribution to learning throughout the semester. Each student has points equal to ten times the number of teammates s/he has and must allocated those points to his/her teammates such that at least one score is differentiated, thus not everyone can receive a ten. Evaluations range from 0 to 25, but typically fall between 7 and 13 inclusively, with 87% of all evaluations falling in this range. The lowest average any individual who completed the class received was 0.8 and the highest was 14.7. Only seven students received average peer evaluations below 7.4 and only four received averages greater than 13. Team cohesiveness is measured as the negative of the variance of all the individual peer scores for each team.

Diversity indices are created for gender and class level using an inverse Herfindahl index such that diversity is measured as the inverse of the sum of squared proportions. For gender, this index can range from 1 for a team with no gender diversity to 2 for a team that is 50 percent male and 50 percent female. The class level diversity index can range from 1 for a team for which all members are at the same class level, to 4 for a team evenly balanced in terms of freshman, sophomores, juniors, and seniors. Diversity of student ability or effort as reflected by GPA is measured in two ways, first as the difference between the GPA of the highest individual on the team and the team average GPA and second as the difference between the team average GPA and the lowest individual GPA. This total range of GPA for each team was also considered but was not significant in any models so was omitted. What is of interest with the first of these GPA diversity measures is to determine if it matters not just how academically capable a team member is, but if it also matters how much "smarter" this person is than the team average. For example, it might be expected that a student with a 3.9 GPA would add more to a team with an average GPA of 2.8 than to a team with an average GPA 3.4. The second variable is similar, but in the opposite direction, attempting to determine if a low GPA student pulls down a team more, or creates more disharmony on a team, the further below the team average that student is. These diversity measures are summarized in Table 2.

Empirical Model:

Variance in peer evaluations is modeled as a function of surface-level and deep-level variables, as well as team size and class size. In general, the relationship between diversity and team cohesiveness, or lack thereof as reflected by a greater variance in peer evaluations, is ambiguous, with theory suggesting that diversity brings more experiences and different perspectives to the table that can contribute to success, but also creates more differences and possible lack of harmony. Gender and class level diversity indices are used to measure these aspects of diversity. The majors variable and out-of-state were also tested, but found to be insignificant. As there were a variety of different majors on all teams and so few in applied economics, it is likely that this variable was not actually measuring anything of theoretical relevance. While combining out-of-state and in-state students provides increased diversity, it is also possible that in-state students from different parts of the state can provide enough diversity

of perspective, given the geographic variety in South Carolina. Thus the actual diversity measured by this variable is also questionable and both this and the majors variable were omitted from further study.

In addition to surface-level diversity, variation in academic ability and/or effort as reflected in GPA is expected to influence peer evaluations. A high GPA student would be expected to be a stronger than average contributor to a team, perhaps garnering significantly higher peer evaluations as a result. On the other hand, some high GPA individuals are initially less enthusiastic about TBL (Espey 2010), thus might be perceived as aloof or as less than ideal teammates by their peers. It is also possible that success stimulated by one strong student on the team makes all the students on the team more enthusiastic about working hard for continued success, increasing team cohesiveness and reducing variation in peer evaluations. This effect is also measured by the difference between the highest GPA on the team and the team average GPA, rather than just the level of the highest GPA. Similarly, the effect of variation in academic ability on the low end is estimated using either the lowest individual GPA on the team or the difference between the team average GPA and the lowest GPA. These variables are expected to be positive, measure the negative influence of students who tend to skip class more often and are less prepared and participate less actively when they do come to class, thus increasing the variation in peer evaluations within a team.

Team success on RATs and in-class team activities are also both modeled as a function of surface-level and deep-level variables, as well as team size and class size. Gender measured as the percent of the team that is female as well as gender diversity are estimated. The influence of class level diversity is estimated with several measures, including percentages at each level, percent upperclassmen, and a using the diversity index. The influence of grades is also estimated in several ways: the highest individual grade point average (GPA) of students on the team, the lowest GPA on the team, the average GPA of all team members, the difference between the highest GPA on the team and the team average, and the difference between the team average and the lowest GPA on the team. Team cohesiveness is measured using the negative of the variance in the peer evaluation scores for each team, such that greater variance corresponds to less cohesiveness.

The TBL literature recommends teams between 5 and 7 in size, suggesting that there are not enough resources for smaller teams to address complex application exercises while it is harder for all students to contribute substantially in larger teams. Given that all but one team is within the range of TBL recommended team sizes, the team size variable is not expected to be significant, yet finding this empirically is also valuable in that it would support the TBL literature. Class size is also not expected to be a significant factor influencing either peer evaluations or team success, as it is interactions within the teams, rather than between teams, that would be expected to matter most. However, since many potential adopters of TBL question their ability to use it in large or even medium size classes, this variable is included in the analysis.

Empirical Results:

Variance in Peer Evaluations:

First, the results of two models estimated to explain the variance in peer evaluation scores are shown in Table 3. No measures of gender, class level, or gender or class level diversity were

significant in explaining this variance, including percent female, individual class level percentages, or the diversity indices shown in Table 3. One model was estimated using the highest and lowest individual GPA for each team and one using the difference between each of these values and the team average GPA. Neither GPA variable was significant in the first model, but both were significant in the second. As expected, the greater the difference between the lowest individual on the team and the team average GPA, the greater the variance in peer evaluations, usually as a result of significantly lower scores for this lower GPA individual. In contrast, the greater the difference between the highest individual GPA and the team average GPA, the smaller the variance in peer evaluations. Whether this is due to the high GPA individual not being perceived as any better of a contributor than any other students or because possible greater success related to this individual made all team members work better together cannot be determined from these results however.

While team size was not statistically significant, class size was. However, when a dummy variable is included for the one large class of 103 students and the class size variable is restricted to measuring just the other classes, that ranged in size from 19 to 60, class size is not significant but the large class dummy variable is. This may be due to the large size of the class, the different physical environment of the classroom (in a tiered lecture hall with immovable seats), or some other feature of the class cannot be determined at this point. This class was the first agricultural economics course taught at the university using TBL and was also the class with the most freshmen, at 51% compared to an average of 20% in the other classes, with most of them first semester freshmen as this was a fall semester class. In addition to adjustment difficulties many freshman face, this also meant GPA data was available for less than half the classes, thus these resources may not have been as evenly distributed as in other classes.

Team Success:

One model was estimated for the RATs and another for the in-class application activities. The overall team score was a weighted function of these so was not estimated separately. The only variable estimated to be statistically significant in explaining RAT scores was the highest individual GPA on the team. The RATs are intended to test at the basic conceptual level and do not involve application or calculations, as they are given before any lecture or discussion of the material takes place in the class. As such, it is not surprising that the most academically successful students have the most significant impact on the overall team RAT. Interestingly, team cohesiveness does not significantly impact scores, but as students work in the same teams throughout the semester, they likely learn quickly who on the team does well on the RATs and typically go with the majority vote anyway when deciding on team answers for the RATs.

Factors influencing the overall grade on the 12 to 13 application exercises were also estimated. Neither the gender nor the class level diversity indices were significant, but the percentage of the team that was female positively influenced team success. A squared term for females was also included to determine if there was a diminishing impact but it was not significant. This result corresponds with some of the sociological literature that finds that females invite more discussion and openness on teams, factors that are important for many of the application exercises that involve choosing “the best” among a set of options and those involving justification of answers when more than one choice might be correct.

The influence of class level was estimated in a variety of ways: including the class level diversity index, a dummy variable for three of the class levels, and a dummy variable for upper classmen. Only the dummy variable for juniors was found to be significant and somewhat

surprisingly had a negative influence on team success. This is perhaps that the juniors in the class tend to have below average grades, not meeting with success in some of their science courses and delay taking economics. Being upperclassmen, these students may be more self-confident, particularly in comparison to first semester freshmen, and other students may defer to their judgment.

In terms of success on application activities, both team cohesiveness and an academically strong student on the team contributed positively. Neither team size nor class size were significant. This latter result is interesting in light of the earlier finding that the large class had more variability in the peer evaluations; so while the teams in this class were less cohesive on average, they were able to overcome this and be as successful on team activities as teams in other classes. The lack of significance of team size suggests that teams of five to seven do not differ significantly in terms of the resource availability, but the limited range of team sizes in this study precludes making conclusions about teams larger or smaller than this range.

Conclusions:

Awareness of characteristics of team composition that contribute to team success can help faculty members more carefully design teams to enhance outcomes and learning. For example, more effective teams will have members with complementary skills, but effectiveness also appears to be enhanced by team cohesiveness, thus faculty should be attuned to divisions or conflicts on teams and attempt to alleviate such problems to the extent possible. Explaining to students the value beyond the classroom of learning to work in teams helps students see the bigger picture. Businesses use teams for many projects; the literature addressing the team composition in organizations often addresses who *not* to put on a team, yet in the classroom, everyone must participate if an instructor chooses to use teams or groups. Knowing what team components matter and what don't, in terms of team success, can help instructors focus energy and attention on the appropriate variables in designing teams and working to maximize team performance. If improving team performance in the classroom can help students better learn to work in teams, it could ultimately improving their employability and work productivity as well.

References:

Bell, S.T. (2007). "Deep-Level Composition Variables as Predictors of Team Performance: A Meta-Analysis", *Journal of Applied Psychology*, 92(3): 596-615.

Bunderson, J.S. and K.M. Sutcliffe (2002). "Comparing Alternative Conceptualizations of Functional Diversity on Management Teams: Process and Performance Effects," *Academy of Management Journal*, 45: 875-893.

Campion, M.A., G.J. Medsker, and A.C. Higgs (1993). "Relations Between Work Groups Characteristics and Effectiveness: Implications for Designing Effective Work Groups," *Personal Psychology*, 46: 823-850.

Cossé, T.J., D.N. Ashworth, and T.M. Weisenberger (1999). "The Effects of Team Size in a Marketing Simulation," *Journal of Marketing Theory and Practice*, 7(Summer): 98-106.

Cummings, T.G. (1981). "Designing Effective Work Groups," *In Handbook of Organizational Design, Vol. 2*, edited by P.C. Nystrom and W.H. Starbuck. New York: Oxford University Press.

Dawe, H.C. (1934). "The Influence of the Size of Kindergarten Group on Performance," *Child Development* 5: 295-303.

Deeter-Schmeltz, D.R., K.N. Kennedy, and R.P. Ramsey (2002). "Enriching Our Understanding of Student Team Effectiveness", *Journal of Marketing Education*, 24(2): 114-124.

Denton, H.G. (1994). Simulating Design in the World of Industry and Commerce: Observations from a Series of Case Studies in the United Kingdom," *Journal of Technology Education*, 6(1): 1045-64.

Devine, D.J. and J.L. Phillips (2001). "Do Smarter Teams Do Better: A Meta-analysis of Cognitive Ability and Team Performance," *Small Group Research*, 30: 678-711.

Dommeyer, C.J. (1986). "A Comparison of the Individual Proposal and the Team Project in the Marketing Research Course," *Journal of Marketing Education*, 8(Spring): 30-38.

Espey, M. (2010), "Valuing Teams: What Influences Student Attitudes?" *NACTA Journal*, 54(1):31-40.

Hackman, J. R. (1987). "The Design of Work Teams," In *Handbook of Organizational Behavior*, edited by J.W. Lorsch. Englewood Cliffs, NJ: Prentice Hall.

McGrath, J.E. (1964). *Social Psychology: A Brief Introduction*, New York: Holt.

Meyer, J. (1994). "Teaching teams through teams in communication courses: Letting structuration happen." Paper presented at the annual meeting of the Speech Communication Association, New Orleans.

Michaelsen, L.K., A.B. Knight, and L.D. Fink (2002). Team-Based Learning: A Transformative Use of Small Groups in College Teaching, Praeger Publishers, Westport, CT.

Michaelsen, L.K. and M. Sweet (2008). "Team-based learning: Small group learning's next big step," *New Directions in Teaching and Learning*, No. 116.

Miller, N.E. (1951). *The effect of group size on decision-making discussions*, Ph.D. dissertation, University of Michigan, Ann Arbor.

National Association of Colleges and Employers, http://www.naceweb.org/

Nichols (2002). "The Effects of Cooperative Learning on Student Achievement and Motivation in a High School Geometry Class," *Contemporary Educational Psychology*, 21(4): 467–476

Pelled, L.H., K.M. Eisenhart, and K.R. Xin (1999). "Exploring the black box: an analysis of work group diversity, conflict, and performance," *Administrative Science Quarterly*, 44(1): 1-28.

Salas, E., T.L. Dickinson, S.A. Converse, and S.I. Tannenbaum (1992). "Toward an understanding of team performance and training." In *Teams: Their Training and Performance*, edited by R.W. Swezey and E. Salas, 3-30. Norwood, NJ: Ablex.

Stewart, G.L. (2006). "A Meta-Analytic Review of Relationships Between Team Design Features and Team Performance." *Journal of Management* 32: 29-55.

Sweet, M. and L.K. Michaelsen (2011). *Team-Based Learning in the Social Sciences and Humanities: Group Work That Works to Generate Critical Thinking and Engagement*, Stylus Publishing, LLC.

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

Webber, S. S., and L.M. Donahue (2001). "Impact of highly and less job-related diversity on work group cohesion and performance: A meta-analysis." *Journal of Management*, 27: 141-162.

Yeatts, D.E. and Hyten, C. (1998). High-Performing Self-Managed Work Teams: A Comparison of Theory and Practice. *Thousand Oaks, CA: Sage.*

Table 1: Summary Statistics

	Mean	Standard deviation	Minimum	Maximum
Surface Level Variables:				
Female (%)	42.8	21.1	0	100
Class Level(%):				
Freshmen	27.6	20.5	0	71.4
Sophomore	39.2	16.5	0	80.0
Junior	21.7	18.9	0	80.0
Senior	11.4	12.4	0	50.0
Major (%)	81.0		0	100.0
Out-of-state (%)	24.0	18.5	0	83.3
Deep Level Variables:				
Max gpa	3.69	0.26	2.70	4.00
Average gpa	2.92	0.18	2.35	3.40
Lowest gpa	2.09	0.38	1.00	2.86
Variance in peer evaluations	4.47	6.71	0.41	53.68
Environmental Variables:				
Team size	6.04		4.0	7.0
Class size	46.6		19.0	103.0

Table 2: Diversity measures

	Mean	Standard deviation	Minimum	Maximum
Surface Indices:				
Gender	1.72	0.27	1.00	2.00
Class Level	2.57	0.58	1.47	3.77
Deep Variables:				
Top individual GPA minus team average GPA	0.77	0.22	0.23	1.44
Team average GPA minus lowest individual GPA	0.83	0.27	0.27	1.63

Table 3: Results for Variance in Peer Evaluations

Team Variable	Model 1	Model 2
Surface Level:		
Gender diversity	0.66 (0.26)	-0.39 (-0.16)
Class level diversity	-1.36 (-1.11)	-1.59 (-1.33)
Deep Level:		
Highest individual GPA on team	-2.24 (-0.88)	
Lowest individual GPA on team	-2.88 (-1.62)	
Top individual GPA minus team average GPA		-6.802** (-2.01)
Team average GPA minus lowest individual GPA		7.927*** (2.77)
Environmental:		
Team size	0.71 (0.65)	0.76 (0.71)
Class size for up to 60 students	0.02 (0.29)	0.02 (0. 24)
Dummy for 103 student class	7.70** (2.12)	7.87** (2.22)
R-squared	0.23	0.27

Table 4: Results for Team Performance

Team Variable	RAT Scores	Activity Scores
Surface Level:		
Percent Female	0.0099 (0.70)	0.046** (2.08)
Percent Juniors	-0.0017 (-0.10)	-0.046* (-1.84)
Deep Level:		
Highest GPA	0.030*** (2.79)	0.04** (2.41)
Team Cohesiveness	0.00012 (0.26)	0.0015** (2.20)
Environmental:		
Team size	0.0021 (0.51)	0.00074 (0.012)
Class size	-0.000012 (-0.10)	-0.00012 (-0.65)
R-squared	0.10	0.19