University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Theses, Dissertations, and Student Research: Department of Psychology

Psychology, Department of

8-1-2012

MAKING A LARGE CLASS FEEL SMALL USING SOCIAL PSYCHOLOGY: INTRODUCING TEAMS TO IMPROVE PERFORMANCE AND LEARNING IN A LARGE-ENROLLMENT COURSE

Bethany Johnson
University of Nebraska-Lincoln, bethc.johnson@yahoo.com

Follow this and additional works at: http://digitalcommons.unl.edu/psychdiss

Part of the Psychiatry and Psychology Commons, and the Social Psychology Commons

Johnson, Bethany, "MAKING A LARGE CLASS FEEL SMALL USING SOCIAL PSYCHOLOGY: INTRODUCING TEAMS TO IMPROVE PERFORMANCE AND LEARNING IN A LARGE-ENROLLMENT COURSE" (2012). *Theses, Dissertations, and Student Research: Department of Psychology.* Paper 49.
http://digitalcommons.unl.edu/psychdiss/49

This Article is brought to you for free and open access by the Psychology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Theses, Dissertations, and Student Research: Department of Psychology by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

MAKING A LARGE CLASS FEEL SMALL USING SOCIAL PSYCHOLOGY: INTRODUCING TEAMS TO IMPROVE PERFORMANCE AND LEARNING IN A LARGE-ENROLLMENT COURSE

by

Bethany C. Johnson

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

Major: Psychology

(Social Psychology)

Under the Supervision of Professors Elizabeth Crockett and Marc Kiviniemi

Lincoln, Nebraska

August, 2012

MAKING A LARGE CLASS FEEL SMALL USING SOCIAL PSYCHOLOGY: INTRODUCING TEAMS TO IMPROVE PERFORMANCE AND LEARNING IN A LARGE-ENROLLMENT COURSE

Bethany C. Johnson, Ph.D.

University of Nebraska, 2012

Advisors: Marc Kiviniemi and Elizabeth Crockett

Large-enrollment lecture-based classes are increasingly common in higher education. As an alternative approach, active learning methods are meant to develop academic skills and improve understanding of course content. Group work is an effective form of active learning, but students typically despise it. Social psychological small group theory can inform teachers about the characteristics of small groups that influence their capability to improve learning, so that teachers can design more effectual group work for their classes. This study examined what effect introducing permanent teams into a large enrollment class had on students' sense of classroom community and their learning outcomes, using both exam performance and writing scores as objective measurements. This study employed a non-equivalent control group quasi-experimental design, and used the first of four sequential semesters of the same course as a baseline comparison group. I hypothesized that students would report a stronger sense of community in the semesters including teams, and that learning outcomes, as reflected in exam scores and grades on the writing assignment, would improve as well. The teaching innovation did not produce the desired and predicted outcomes, but the results still constitute progress toward developing a successful intervention. Limitations to the

present study are described in terms of recommendations for future research on the strategic integration of the scholarship of teaching and learning and social psychology. With this approach in place, teachers can begin to establish best practices for group work in large-enrollment classes.

Keywords: Social Psychology, teams, large-enrollment classroom, small groups, teaching methods, classroom community

Acknowledgements

I have benefited greatly from help, support, and encouragement from my friends and family, as well as my advisor, supervisory committee, and colleagues. It would be impossible to list out the names in the space available here of all the people who deserve recognition, so please know that you are listed in my heart. This would not have been possible without you all. Thank you. Special thanks to Andrea Burnham, who dedicated hours to helping me polish the final draft.

Table of Contents

Abstract ……..ii

Acknowledgements ……iii

Table of Contents ……...iv

List of Multimedia Objects …….v

Chapter 1: Introduction ……..1

Applying Social Psychology to Improve Classroom Interventions ……...9

Overview of the Present Study …….…….…..…23

Hypotheses ……..……...….24

Chapter 2: Method …….…….……25

Materials and Procedures ……29

Chapter 3: Analysis ……..…..…….……44

Chapter 4: Results …….…….……61

Chapter 5: Discussion …….……....79

Implications and Conclusions …….……91

References …….…….….95

Appendices …….…….100

List of Multimedia Objects

Table 3.1: Logistic Regression Values for Binary Covariates across Semester …….46

Table 3.2: Original Descriptive Statistics on the Variables that Comprised the Propensity Score Covariates …….48

Table 3.3: Propensity Scores by Quintile per Semester …….52

Table 3.4: Results of Ordinal Logistic Regression of Semester and Quintiles on Students' Year in School ……....54

Table 3.5: New Descriptive Statistics on the Propensity Score Covariates after Culling Mismatched Cases ……..56

Table 3.6: Logistic Regression Values for Binary Major across Semester ……57

Table 3.7: Original Descriptive Statistics for the Course Evaluation Covariates that Comprised the Propensity Scores ……58

Table 3.8: Propensity Scores by Quintile per Semester for Course Evaluations ……59

Table 3.9: Results of Ordinal Logistic Regression of Semester and Quintiles on Students' Year in School ……60

Table 3.10: New Descriptive Statistics for the Propensity Score Covariates After Culling Mismatches ……..60

Table 4.1: Classroom Community Scale Scores by Semester, Including Propensity Score Adjustment ……..63

Table 4.2: Classroom Community Scale Scores by Semester, Analyzed Without Propensity Score Adjustment ……..66

Table 4.3: Repeated Exams Within-Subjects Contrast F-Test Results ……...69

Figure 4.1: Performance Across Exams by Semester …….70

Table 4.4: Means of Exams by Semester ……71

Table 4.5: Course Recommendation and Activities' Helpfulness by Semester …….74

Table 4.6: Logistic Regression Values for Course Recommendation by Semester ……75

Table 4.7: Discussion Participation Grade by Semester …….77

Table 4.8: Means of Students' Evaluation of Writing Assignment by Semester ……78

Chapter 1: Introduction

Group work in the classroom is a valuable teaching method. Groups not only allow collaboration and cooperation among students, who can use their peers to assimilate the information provided by the teacher or the course materials, but group work can also teach students how to work productively with others, which is an important skill in academic or professional settings (Barfield, 2003; Cohn, 1999; Johnson, Johnson, & Stanne, 2000; Lightner, Bober, & Willi, 2007; Monk-Turner & Payne, 2005; Woo & Reeves, 2007). Unfortunately, badly designed group activities can produce antipathy in students and hinder learning (Barfield, 2003; King & Behnke, 2005). Social psychology includes much research and theory regarding group composition, the dynamics between members, how groups function differently from individuals, and how groups can improve individuals' performance. Using this information, teachers can design group activities that maximize their potential to improve learning outcomes.

The present study examined the effect of including permanent teams in a large-enrollment intermediate level course as an ancillary course component on students' sense of community in the class and their learning of the material. The teams were designed to create a context within which the students could autonomously earn their own grade while still experiencing the connection to classmates, similar to that which is possible in small classes. This improved connection and interaction with their classmates should correspond with improved learning as reflected in their grades, and have a visible impact on the performance of a semester-long group assignment. Therefore, my main goal was to create a social and academic support structure that would improve students' subjective feelings of connection to their classmates, in order to improve their learning of course

material. Research on the effect of class size on students' learning has produced mixed results, though researchers surmise that its inconclusiveness is due to a failure of studies to account for the myriad factors that interact with class size to impact learning (Chapman & Ludlow, 2010; Pedder, 2006). Pedder concludes that class size in isolation does not have a clearly positive or negative effect (i.e., large classes are not clearly worse or better for learning than small classes), but that the quality of student and teacher interaction is what moderates the relationship between enrollment numbers and learning outcomes. For example, the amount of time teachers can offer personalized feedback decreases as class size increases. Furthermore, the opportunities to incorporate active learning course components that develop students' other academic skills are limited the larger the class becomes (Exeter et al., 2010; Pedder, 2006).

As enrollments in universities are rising to keep up with the increasing cost of education, the prevalence of large enrollment classes is also escalating (Chapman & Ludlow, 2010; Mulryan-Kyne, 2010). Some researchers think that change might adversely affect retention at the university level and attrition at the course level (Mulryan-Kyne, 2010). Whether or not the size of the classes' enrollments is an influence on retention directly, the uncontested fact is that the larger the class, the more strain it puts on the instructor to effectively and efficiently reach his or her teaching objectives (Chapman & Ludlow, 2010; Exeter et al., 2010). The review below will address both the teachers' role in and the students' perspective on the influence of groups in the classroom, and the benefits of including active learning components in the context of social psychology.

Social Psychology of Groups

Social psychology goes a long way toward explaining how groups work, and what effect groups have on performance outcomes, all of which is particularly relevant in educational settings. According to the definition held by most social psychologists, groups are collections of two or more people who are interdependent, have a common goal, and interact with each other directly. Members of a group are aware of their collective identity, and the group exists for a meaningful length of time. Meaningful in this situation refers to personal significance, as opposed to a more objective evaluation of the duration of association (Levine & Moreland, 1998). A group is socially integrated when people start acting more like group members than individuals. *Entitativity* is the term for the point at which a collection of people becomes a group, possessing the goals, characteristics, and behaviors of a single unit (Campbell, 1958, as cited in Pickett, 2001). Research on groups shows that entitativity exists along a continuum, and how strongly a group identifies as a single entity determines the members' behavior and cognitions, as well as influences how others perceive the group members. A group is high in entitativity when the members are similar to each other, when they share common goals, and when their outcomes are interdependent. Group members will classify people not in the group as outsiders and compete with or discriminate against them, often unintentionally (Tajfel, Billig, Bundy, & Flament, 1971). Outsiders will implicitly perceive group members to be closer to each other physically and psychologically than people who are not in a group (Pickett, 2001). An *in-group* is a group to which a person belongs (i.e., "Us"); an *out-group* is a group of which a person is not a member (i.e., "Them"). The present study particularly focused on social psychology's official definition of groups to design a best-practices approach to designing an intervention. The

features of a group that can be manipulated to achieve entitativity in small groups in the classroom are collective identity, superordinate goals, frequent interaction, interdependence, and out-group competition.

According to Levine and Moreland (1998), groups have a collective identity, and exist for a meaningful (according to group members) length of time. In academia, semesters provide predictable and widely-acknowledged units of time during which students belong to a class. Students in a class can develop a collective identity if they think of themselves as members of that particular class, and are recognized as such by out-group members. For small groups within a class, however, collective identity can be made more explicit by creating the markers of a collective identity: a group name, a shared motto, and a mascot. These three elements have been used successfully to create collective identities in classroom groups before (see Johnson & Johnson, 1994; Lancy & Rhees, 1994). Another element necessary for a group identity to exist is an out-group – people who do *not* belong to the collective. Groups within the classroom can emphasize this separation by creating competition or social comparison among other student groups in the classroom (Mullen & Copper, 1994).

A classic study by Muzafer Sherif (1956) demonstrated creation of collective identity and the role of superordinate goals in reducing conflict between groups. In a study designed as a summer camp for boys, Sherif and his colleagues chose boys from similar demographic backgrounds who did not previously know each other, and from the first bus ride to the camp onwards, he observed how the boys made friends and connections. During the time they were there, the boys were split into two groups, who spontaneously named themselves and formed collective identities. Within the groups, the

boys ascribed roles to each member, regarding delegation of chores, as well as social roles such as leaders. The two groups then became competitive and even aggressive with each other, taunting and baiting boys that had been their friends only a day before on the bus. Once Sherif and his colleagues saw that the groups were sharply delineated, they attempted to join the two groups back together and alleviate the animosity that had developed. They devised several "problems" that the groups had to work together to solve; for example, they sabotaged the water main leading from the tower to the camp, and asked the boys to all help find the problem and fix it. Another situation required the boys to collectively choose a movie and then pool their money to rent it for the whole camp to watch together. With these activities, the researchers effectively reintegrated the boys back into one group. Sherif concluded that cooperation was necessary to reduce inter-group animosity and competition, and cooperative work on a common goal facilitated the development of entitativity.

Goals and Interdependence

Groups need a common goal or purpose to function. A goal is specifically an objective to obtain within a particular amount of time (Locke & Latham, 2002). Goals serve to direct and organize efforts, are energizing, promote perseverance, and activate associated knowledge and tactics (Locke & Latham, 2002). Identification of common group goals solidifies the group's sense of in-group identity (Gaertner et al., 2000). A group is able to pool its members' skills and knowledge to meet the group's goals. Competition between group members' personal objectives makes it more difficult for a group to achieve its goals; when a group's members' goals are in concert, performance quality improves (Locke & Latham, 20026). If feedback regarding the group's progress

toward their goal is given in terms of members' collective effort, as opposed to evaluations of individual members' contributions, members focus on team goals instead of personal goals (Locke & Latham, 20026). Tajfel and his colleagues (1971) pointed out that the expectation of future interaction with group members increased the strength of the bonds within the group, even when there was no out-group in competition with the group to increase its solidarity.

In the classroom, students interact every day that the class meets, and their personal goals are ostensibly to pass the class (e.g., performance goals) and learn the material (e.g., mastery goals). Hijzen, Boekaerts, and Vedder (2006) note that students also have entertainment, social support and belonging, and self-exploration goals that all converge in the classroom as well. Hijzen and her colleagues collected data on nearly 2000 students regarding their stated goals in the classroom, their opinion of the cooperative learning climate of their classroom, and their motivation to either master the material or merely perform well enough to earn a satisfactory grade. Students stated that their primary goals in the classroom were to learn the material, but a close second was their pursuit of social and belonging goals. Students with strong social goals were the most satisfied with cooperative learning in the classroom.

Students' social and entertainment goals in the classroom are also sources of interdependence among their classmates. Aronson and Bridgeman (1979) point out that students in classrooms might work toward shared goals and still be selfishly motivated to achieve their own personal goals – the work with their classmates just achieves shared goals along the way. This dual purpose is not problematic, just a realistic description of the complexity of the classroom environment. Explicit interdependence and shared goals

in the classroom can occur in group work and activities that require students to come together. The implicit interdependence of the class is evinced through students' dependence on those willing to speak up and ask questions and the collective benefit of hearing the answers to students' questions in class. Furthermore, students' performance on assessments would create the distribution used in a curved grading system that affects everyone, or in the occasion of test items being discarded because everyone struggles with them.

Classroom Community

Classes should be considered groups, in the social psychological sense (Hart, 1995). The idea that whole classes can be classified as groups already exists in the literature, but they are instead sometimes called communities of practice or learning communities (Dawson, 2006; Rovai, 2002a). *Classroom community* as a psychological construct is comprised of two dimensions: a social community and a learning community. Classroom communities, in the same way as social communities, consist of people in proximity to each other who share interests and history, are interdependent, and feel a sense of belonging and cohesion (Dawson, 2006; Rovai, 2001, 2002a; Rovai & Wighting, 2005; Summers, Beretvas, Svinicki, & Gorin, 2005; Summers & Svinicki, 2007; Urdan & Schoenfelder, 2006). A learning community is organized around shared and internalized education norms and values, and members expect to have their learning goals met through the community (Rovai & Wighting, 2005). Researchers typically define classroom community as a combination of affective components of trust, care, and safety, in the specific context of an educational setting, with shared emphasis on learning and understanding, and existing for a fixed length of time (Dawson, 2006; Rovai, 2001,

2002a; Summers & Svinicki, 2007). The qualities included in this definition were derived from multiple studies developing and testing measurements of students' sense of community in classrooms, using factor analysis to identify underlying constructs and extensive replications to establish reliability.

Classroom community has a strong positive relationship with students' motivation and performance in the class (McKinney, McKinney, Franuik, & Schweitzer, 2006). The stronger the students' sense of community in their class is, the better they perform on class work and the more they report liking the subject (McKinney et al., 2006, Rovai, 2001). This beneficial effect is modified by the nature of the task the class is attempting to accomplish; if the community's primary focus is task completion, as opposed to social interaction, members who have internalized that as the purpose of the community will benefit (Rovai, 2001; see Evans & Dion, 1991, and Langfred, 1998, for corroboration).

The atmosphere of the class, or the classroom climate, reflects the students' sense of community (Fassinger, 2000; Urdan & Schoenfelder, 2006). Climate encompasses everyone's attitudes toward learning, norms for behavior, and patterns of interaction. Classroom climate is reciprocally related to students' interaction; students' interactions with each other and the instructor creates the climate, and students' likelihood and quality of interaction with the class and the instructor is also influenced by the climate. McKinney, McKinney, Franuik, and Schweitzer (2006) conducted a study in which planned classroom community building activities and assignments were incorporated into an undergraduate psychology course. The activities and assignments were tailored to cultivate the specific qualities encompassed by the definition of community – social connectedness, safety, trust, belonging, and active participation in the group. The

researchers found the strongest improvements from the beginning of the semester to the end in the learning measures (i.e., exam performance) for those students who reported the strongest sense of community, and all students reported increased satisfaction with the course. Classrooms do not spontaneously cultivate a sense of community; instead, it must be fostered and shaped, and instructors can encourage its development by emphasizing their care for students' learning and creating a safe and positive environment (McKinney et al., 2006; Summers & Svinicki, 2007).

Applying Social Psychology to Improve Classroom Interventions

Conceptual Foundation of the Present Study

The growing trend of large-enrollment classes is leading to a mismatch of teaching strategies with appropriate content, which results in lower-quality learning. Teachers assert that students in large-enrollment classes are more passive, more anonymous, and less engaged with the material and each other than are students in small classes (Kreie, Headrick, & Steiner, 2007; Messino, Gaither, Bott, & Ritchey, 2007; Summers & Svinicki, 2007). The current zeitgeist in higher education improvement touts the benefits of *active learning* over "traditional" instructive methods such as lecture and seminar (Fox-Cardamone & Rue, 2003; Gray & Madson, 2007; King & Behnke, 2005; Long & Coldren, 2006; Messino et al., 2007). It is not that lectures are inherently bad; lecture format is the most efficient way to deliver large quantities of information to large groups of people (Vesilind, 2000), but it is not the most effective way to teach many types of information or skills. Active learning is the recommended antidote for the shortcomings of large-enrollment classes and indiscriminately applied lecture format (Fox & Rue, 2003; Gray & Madson, 2007; King & Behnke, 2005; Lightner, Bober, &

Willi, 2007; Long & Coldren, 2006; Machemer & Crawford, 2007; Messino et al., 2007; Smith, 1996). Active learning (sometimes called student-centered learning) allows students to better reflect, evaluate, synthesize, and communicate information with their classmates, in stark contrast to lectures where the teacher delivers information to a passive audience (Lightner et al., 2007; Machemer & Crawford, 2007; Smith, 1996). It also distributes the onus for learning more evenly across students and teachers. Many active learning methods group students together within the class (Healey & Matthews, 1996). When done properly, group work can be one of the most effective types of active learning (see the meta-analysis by Johnson, Johnson, & Stanne, 2000), because students use interaction with their peers to help them assimilate and apply the information given by the teacher (Healey & Matthews, 1996; Summers, Beretvas, Svinicki, & Gorin, 2005). In addition, students additionally benefit from group work by learning how to work effectively with others, as well as practicing expressing themselves and solving problems cooperatively (Lightner et al., 2007; Machemer & Crawford, 2007; Smith, 1996). Incorporating active learning course components into large classes can be more challenging for teachers than in small classes (McKinney & Graham-Buxton, 1993; Messino et al., 2007). In practice, classrooms become learning environments that allow students to energetically engage course content through problem-solving exercises, informal small groups, demonstrations and simulations, and other activities, with the focus on application of the principles to real life (Johnson, Johnson, & Smith, 1991).

Incorporating groups into a large-enrollment course using recommendations from social psychology is generally predicted to foster a sense of classroom community where students can benefit from scholarly interaction (Summers, Beretvas, Svinicki, & Gorin,

2005; Summers & Svinicki, 2007). Summers and Svinicki conducted a large study of multiple instructors' undergraduate courses, and investigated the relationship between learning goals, the use of interactive learning in classrooms, and students' sense of classroom community. Their methods included multiple quantitative measures of motivation to learn, sense of classroom community, and social interdependence. They found that, when comparing traditional lecture-based classes to classes using cooperative learning methods, students in the cooperative classes reported stronger mastery motivation (as opposed to performance-oriented motivation), and developed a stronger sense of classroom community. Their rigorous analysis of the classroom data offers strong support to the supposition that the sense of community in a classroom and students' learning are intertwined. What characteristics should a teacher address when designing groups and group work in order to improve the connection between students, avoid social loafing, encourage collaboration and cooperation, and increase student learning, while avoiding the characteristics that make students hate group work and sap groups' effectiveness? The purpose of the present study is to examine the effectiveness of having a group of classmates for each student to use as a learning resource and get to know better than they otherwise might in a regular large-enrollment course. Can the social connections and resources provided by groups that are not required as a graded course component improve students' performance and experience in groups that *are* part of the graded course requirements? The elements of group work examined in the present study are a step toward synthesizing the best practices suggested by social psychology for group design.

Steps toward Best Practices in Group Design

Collaborative or cooperative learning consists of two or more students working together to achieve understanding, solve a problem, or create a product, by joint intellectual effort (Delucchi, 2006). As widespread as cooperative learning pedagogy is in higher education classrooms, Delucchi points out that many of the articles describing its value are based on anecdotal evidence instead of quantitative assessments of learning outcomes. Delucchi systematically evaluated the effect of collaborative learning assignments on exam scores in an undergraduate statistics course. He incorporated multiple group projects into the course and analyzed the predictive value of the project grades on exam scores using data from eight sections of the course. Students chose their own groups and formed new groupings for each subsequent project. While he did not find conclusive evidence that the group projects positively affected exam scores early in the semester, toward the end of the semester, new projects had a more significant positive effect on final exams (Delucchi, 2006). My assessment of his study leads me to believe his results were likely confounded by the fact that the groups in his class were different for each project, and because they were self-selected, it is probable that students were more successful at choosing "good" groupmates after experiencing unsatisfactory outcomes with others. A better strategy might instead be to maintain permanent groups across the semester and work to equip the group members with the skills or resources necessary to be increasingly more effective as a collective. Furthermore, any test of the group's effect on its members' learning would require that groups were permanent across the semester, so that any effect of the groups would not be confounded with different groupings of students.

Lightener, Bober, and Willi (2007) collected studies that evaluated group work and active learning methods in order to see how such strategies are typically assessed, and found the majority of researchers used objective learning outcomes (i.e., exam grades) to evaluate their benefit. Lightner and her colleagues then conducted a study testing their idea to use active, collaborative learning to practice problem-based activities in a graduate level accounting class. They focused on measuring the group processes rather than students' performance. The researchers collected data on observed student interaction patterns and surveyed students on their attitudes toward the class material, each other, the group organization, and the instructor. The researchers found that students reported liking working with their classmates for the most part (although a lecture format was still preferable to many), and the students generally had positive opinions about their group mates. Lightner and her colleagues' study would seem to offer evidence that students' subjective opinions of group work are positive, which is an important element to consider when creating a classroom climate that is conducive to learning. To the extent that group work varies, the study is heartening but not definitive support for the potential beneficial influence of group work on social climate in a generic classroom. Lightner and her colleagues' study was conducted on a graduate level class, which tend to have smaller enrollments and higher caliber students than do many undergraduate classes. It is my opinion that their positive outcome was heavily dependent on that circumstance. Aside from size and aptitude, undergraduate classes are more heterogeneous, in terms of ability, motivation, and training, in contrast to most graduate classes. Students working with similar peers have fewer obstacles to navigate than those working within mixed groups. Lightner's results are not directly comparable

to the present study for the reasons just described. Minimizing the difficulty students feel working within the heterogeneity of a large undergraduate course (often used to fulfill general education requirements) is just one more reason to design a theory-based active learning intervention.

Machemer and Crawford (2007) also attempted to evaluate students' opinions of the group work experience rather than focusing solely on learning or performance outcomes. They surveyed students' ratings of class activities that included traditional lecture-based assignments and cooperative group assignments. Students' opinions showed that they liked working in groups the least of all the activities and, across all of the class activities, valued those activities that helped them with exams the most. Barfield (2003) also conducted a study in which undergraduate students had to complete a group writing assignment as part of his course. He surveyed students' attitudes about group work in general, as well as specifically for that class, and looked for demographic differences in their attitudes toward group work. He found that older students, whether in age or by year in school, held more negative attitudes toward group work, and students with outside work responsibilities found group work frustrating to fit into busy schedules. Barfield's intervention minimally incorporated group activities into the class (fewer than five occasions across the semester), and the researcher worried that spending time on group work took away from time spent covering course material. In my opinion, that attitude suggests that he did not fully believe that integrating group work into the courses was a viable means of learning course content. If group work or other active learning strategies can improve students' learning of course material, a minimalist attempt limits the ability of the social connections formed within groups to influence learning, and it

also communicates to students that active learning is a fringe approach, and the bulk of 'real' teaching is still lecture. To make an illustrative analogy, consider a person trying to quit smoking because they believe quitting smoking is a sure way to improve one's health and well-being. Reducing smoking to weekends only then is an inadequate substitute. While the person would still benefit a bit from the reduction in inhaled carcinogens, the lungs would not truly begin to heal unless all smoking truly stopped. My review of the studies just described led me to conclude that, unfortunately, many instructors' attempts to incorporate active learning into their undergraduate courses fall short. Across the many studies reviewed, it appears that group work attempts often suffer from atheoretical planning, inconsistent application, or the failure of students to invest in the activities. Consequently, students tend to have negative attitudes toward active learning components in courses because of previous experiences with badly designed assignments and activities (Barfield, 2003; Fox-Cardamone & Rue, 2003; King & Behnke, 2005).

Students may not be perceiving or receiving the benefits of group work in the classroom because the group assignments or activities they have experienced have been designed without accounting for what researchers know from social psychological theory about how people and groups function, and how group dynamics can be manipulated to achieve educational goals. Utilization of social psychological principles may aid in the creation of better teaching innovations. In classes where group work is done well, students do tend to report appreciating group work after the fact. For example, some students report liking the opportunity to get to know their classmates after a semester of cooperative work, and welcome pooling their resources (Lancy & Rhees, 1994). One

study reported students liked a “team” approach to learning (Kreie, Headrick, & Steiner, 2007), while another study reported that students claimed that they found group discussions helpful and pleasant (Cannon, 2006). Despite the example of Barfield’s (2003) intervention described above, I would note that even if students do not like doing something, there can be pedagogical value in it. Many students resent group work and assignments in classes because they have had bad experiences with social loafing among their peers and imprecise learning goals behind the group work (Carnes Stevens, 2007).

Combating Social Loafing

A standard complaint that students have about group work in classes is that they have to deal with group mates taking advantage of the group’s ability to conceal a single member’s minimal contribution when the whole group is given one grade (Harkins, 1987; Karau & Williams, 1993; Latane, Williams, & Harkins, 1979). Individuals working alongside others, as well as people working collectively with others, function differently than people working by themselves, in terms of effort and outcome quality. The effect of groups varies depending on the nature of the task. People performing simple tasks will do better in the presence of others than they would alone (Zajonc, 1965). This process is *social facilitation*. Conversely, people’s performance on complex or difficult tasks will suffer in the presence of others, compared to their performance when alone. This outcome is still considered part of the process of social facilitation (Harkins, 1987; Zajonc, 1965). In a related effect, people working together on a task in a group are likely to exert less effort than they would if they were by themselves, which is a phenomenon known as *social loafing* (Harkins, 1987; Karau & Williams, 1993; Latane et al., 1987). Despite their differences, both of these effects share an antecedent (Jackson & Williams,

1985). The ability of others to identify a person's unique contribution to the outcome is one of the driving forces behind both facilitation and loafing (Harkins, 2006). In facilitation, a person's performance will suffer on difficult tasks in part because of evaluation apprehension, which is the anxiety a person feels during appraisal. In loafing, a person working with others, where his or her personal contribution cannot be differentiated from the collective outcome, is relieved of evaluation apprehension, and therefore slackens his or her effort in pursuit of expediency. Therefore, the solution is to make students' individual contributions both evident and explicitly evaluated on their own merits, in addition to the group product assessment (Harkins & Jackson, 1985; Karau & Williams, 1993; Shepperd, 1993). Furthermore, making the task especially engaging, difficult, or enjoyable will discourage loafing (Karau & Williams, 1993). Another strategy for forestalling loafing is to warn them of the possibility, a strategy called *inoculation*, so that they can be vigilant and avoid it (Williams & Karau, 1991).

Part of the difficulty involved in designing and implementing effective group work in the classroom, and especially in large-enrollment classes, is the widespread antipathy students feel toward group work. While the general dislike has been stated above, it is important that teachers understand the intensity of the vitriol in order to plan appropriately to combat it. Numerous studies report that students particularly hate group work because of the time it takes to catch up the "slackers" (King & Behnke, 2005; Kreie, Headrick, & Steiner, 2007), the frustration of students "hitch-hiking" on others' work (Machemer & Crawford, 2007), and the injustice of students who "free-ride" on their classmates' effort (Chapman, Arenson, Carrigan, & Gryckiewicz, 1993; Kerr, 1983; Shepperd, 1993). Good students predominantly suffer the consequences of social loafing,

because they tend to be the ones picking up the slack (Monk-Turner & Payne, 2005). These reports emphasize that one of the primary complaints students have with group work is the issue of social loafing, even though it is rarely identified as such outside of social psychology research. Asked at the beginning of class what elements of a typical class they dislike, group work tops students' list (Fox-Cardamone & Rue, 2003). Students widely report hating group work, dreading it in new classes, and having had negative experiences with it in the past (Barfield, 2003). Researchers coined the term "group hate" in fact, because this pattern of antipathy is so prevalent and vociferously expressed (King & Behnke, 2005). This attitude has likely been the result of their experiences with poorly planned group activities and assignments, and negative interactions with classmates who either take advantage of their hard work, or conversely, take over and dominate the group (Barfield, 2003; Fox-Cardamone & Rue, 2003; King & Behnke, 2005).

Vik (2001) describes many students' misgivings about teamwork in the classroom as springing from difficulties working in groups (e.g., dealing with social loafing and solving interpersonal conflicts), using anecdotal information from years of incorporating cooperative work into her courses. It is not difficult to imagine that, because teachers were once students themselves, they have similar attitudes toward group work themselves. This could be another, less visible impediment to widespread incorporation of group work strategies into courses. To address the standard complaints, Vik recommends using confidential, intermittent peer evaluations that count toward the course or assignment grade to counteract the tendency for students to lessen their own effort toward contributing assuming others will pick up the slack. Furthermore,

evaluating students' individual contributions to the group as well as the group's collective performance allows instructors to assess multiple learning objectives.

Preserving Free Choice

Social psychology can offer suggestions for rectifying many of the typical shortcomings of group work in classrooms. First of all, students primarily worry about the impact of their classmates' behavior on their own grade – some students report feeling as if they lose their individuality and control over their own outcomes when they work in groups (Barfield, 2003; King & Behnke, 2005). As a result, students might display reactance, which means they do the opposite of what they are told to do, in order to preserve a sense of their own free will (Silvia, 2005). One of the potential sources of reactance instructors might seek to avoid is any sense of coercion the students could feel about participating in voluntary group activities outside of class (e.g., study groups). Other strategies can be useful for encouraging student compliance with in-class course component group work. The incentives for participation should be small or count for extra credit in the course, as opposed to required course credit, the loss of which would constitute a punishment for not participating. In addition, small amounts of extra credit might help instructors preserve any existing intrinsic motivation students have to work with their classmates cooperatively, instead of supplanting that interest with a compelling external motivator such as large point value rewards. A classic study in social psychology on what is called the over-justification effect demonstrated that children given rewards for performing an enjoyable task performed that task less in the absence of those expected rewards, compared to children who were not rewarded, or who did not expect to be rewarded for performing the enjoyable task (Lepper, Greene, & Nisbett,

1973). The trick is to not overpower the intrinsic motivation with a strongly compelling external motivator, whether it is positive (i.e., a reward) or negative (i.e., a punishment). The magnitude of the reward does matter, and smaller motivators preserve a person's sense of intrinsic motivation for voluntary behaviors.

Another classic study in social psychology compared the effect of paying a person either one dollar or twenty dollars to lie to another participant about how much they enjoyed a boring, tedious task they had just completed as part of the study (Festinger & Carlsmith, 1959). Asked later about their attitude toward the task, participants who had lied for the small monetary incentive reported actually liking the tedious task, while those paid the large amount reported disliking the task quite a bit. The researchers claimed that this finding demonstrated the power of a small incentive to make people find the reason for doing something inside themselves. Finally, modern empirical evidence suggests that students who are rewarded for working cooperatively, rather than punished for not doing so, will contribute more to a group's effort, especially when they are required to work together repeatedly (Rand, Dreber, Ellingsen, Fudenberg, & Nowak, 2009). Participants in Rand and his colleagues' study played repeated rounds of the public goods game (sometimes called the *commons dilemma*), and the researchers examined the effect of punishing or rewarding participants' contribution to the common good on later games' outcomes. This paradigm, common in laboratory research on cooperation, pits self-interest against group cooperation; participants decide how much of their own resources to contribute to the common pool, which is then divided and redistributed to all players equally. Self-interest would lead a person to contribute nothing, because the pool would still return a share to all players, and the player would maximize his or her own profit.

Cooperation would lead a person to contribute more in order to increase everyone's share. The researchers found that participants' knowing that they were going to have to repeatedly work together with the same people, as well as knowing who contributed to the common good or not in previous games, led to different conclusions than previous research on either anonymous or single game studies on the common good. Participants who rewarded those team members who contributed to the common good had higher payoffs later compared to teams that punished low contributors. These two studies combined demonstrate a good way to approach required group work, in or out of the classroom. Teachers should use small amounts of positive reinforcement to keep students' attitudes about the work positive and preserve their willingness to comply.

Teacher Transparency and Involvement

Hart (1995) described the importance of making the learning goals and benefits of group work known to students. He suggested that students will experience greater group cohesion if the pedagogical benefit of group work is explained, in effect making the teacher's purpose in assigning the work transparent. Hart also recommended that the teacher encourage students' active involvement throughout the duration of the group's interaction, rather than leaving the group to work in isolation. Hijzen, Boekaerts, and Vedder (2006) found that students' perception of the social climate of the classroom significantly predicted their estimation of the benefit of cooperative learning. The authors claimed that cooperative work requires students to have access to each other as resources of both social and academic support; the more they felt their classmates were resources to help them achieve a meaningful educational end, the more they valued the cooperative learning experience. Therefore, the teacher should help make the point

explicit, in order to direct the students' correct perception. Hijzen et al. (2006) recommended that students be monitored while working as groups, both to keep the students on task, but also to communicate teacher support for the students' work in groups. A strategy instructors could use to foster interdependence would be to require students to rely upon each other for things like lecture notes if they missed a day of class (i.e., not posting the lecture slides online), and rewarding students with bonus points for sharing study materials and meeting up to study in groups online or in person. Furthermore, providing ample opportunities for students to communicate with each other informally (e.g., synchronously and asynchronously online) will help them develop a stronger sense of group cohesion (Rovai, 2002a).

McKinney and Graham-Buxton (1993) noticed that many teachers using cooperative learning groups (CLGs) as a teaching strategy were implementing them in the context of small classes, but asserted the CLG concept was equally useful in large-enrollment classes. Their study evaluated the researchers' intervention, which used small groups that were typically self-selected based on seat proximity, and were heterogeneous in terms of students' ability and demographic characteristics. The researchers listed the benefits of the strategy as ranging from opportunities for students to apply course concepts to examples to reducing the anonymity and isolation of large classes, measured through both performance on exams and students comments and ratings on course evaluations. McKinney and Graham-Buxton reduced social loafing in their classroom CLG activities and assignments by having students produce both individual contributions and group products. The assessments of each component were averaged to produce a grade that reflected both the students' personal performance and the collective

performance of the group. McKinney and Graham-Buxton continued the evaluation of the CLGs across four semesters of a large-enrollment introductory sociology course (one baseline semester without the intervention, one semester with informal CLGs, and two semesters with formalized CLGs as a course component), they saw the classes' average final grade rise compared to the semesters without CLG assignments. Students' comments on course evaluations indicated mixed but somewhat more positive reactions to the group work when it was formal, but in the semester with the highest number of formal CLG assignments, students were significantly more interested in taking future sociology courses and found the application exercises more helpful for learning the concepts (McKinney & Graham-Buxton, 1993). This study is the most similar to the present study, but I am extending it further by purposefully incorporated social psychology principles to improve the group work intervention.

Overview of the Present Study

In the present study, I investigated the efficacy of an innovation designed to make a large-enrollment class feel and function more like a small class by introducing permanent small groups of students, conceptualized as *home teams*. These home teams afforded students a smaller set of classmates with whom they could form social connections for sharing academic resources, develop a sense of belonging to a learning community, and have the benefits of group work translate into better learning. The students in the courses including this home team innovation were all arranged into the teams, but were not officially required to participate in the home team activities for course credit; the teams were tangentially related to the course grade using extra credit incentives to encourage participation. In each of the semesters, students completed a

small-group, informal writing assignment that spanned the majority of the semester and took place largely asynchronously online, the purpose of which was to engage in discussion of new social psychology topics each week on an online discussion board. The writing assignment was not part of the team innovation, but during semesters including the teams, the small writing groups were formed out of the teams, and I intended the assignment to build on the existing connections between students to improve the assignment's efficacy as a learning tool. The details of the team innovation design and the group writing assignment specifications are described in the methods section below.

Hypotheses

The present study explored multiple hypotheses, organized thematically below. The first hypothesis purely concerned the predicted effect of the home team innovation on the students' sense of classroom community. I hypothesized that students in classes including teams would report a greater "sense of community" measured by Rovai's *Classroom Community Scale* (CCS; 2002b),

The second set of hypotheses referred to the predicted effect of the home team innovation on students' learning outcomes. I hypothesized that students in classes with teams would earn higher grades on the group writing assignment than students did in the regular format class. Furthermore, I predicted that students in classes including teams would have higher exam scores than students in the normal format class did. Related to the overall exam performance predictions, I predicted students in classes utilizing teams would show more improvement (e.g., rate of change over the course of the semester) across the four unit exams as evidence of the cumulative effect of teams. Finally, I

hypothesized that students with more points for team studying efforts would have higher exam scores than students with no or fewer team studying points.

The third set of hypotheses related to the predicted effect of the team innovation on students' attitudes toward the class. Specifically, I predicted that students in classes including teams would enjoy the course more, which I operationalized as reporting on course evaluations that they would recommend the course to others, compared to students in the normal format class. Furthermore, I hypothesized that students in classes incorporating teams would evaluate the class with a higher rating in response to the question "how good was this course?" than students would in the regular format class. Finally, I predicted students in classes including teams would rate the helpfulness of class activities higher on the course evaluations than students in the regular format class did.

The fourth set of hypotheses focused on the group writing assignment. I predicted that students in classes where the group assignment grade included peer review would participate in the assignment more (i.e., earn a higher score for the discussion participation portion of the assignment grade) than students in the class where there was no peer review element in the assignment grade. The peer review component is one way to make group member interdependence salient to the students. Finally, I predicted that classes from semesters with the peer review element included in the assignment would show a stronger sense of community than classes for whom the writing assignment did not include peer review.

Chapter 2: Method

Overview

The present study evaluated archival records collected in the course of teaching four semesters of an intermediate-level undergraduate social psychology course. The units of analysis were both the student and the class levels. The outcome variables of interest were students' learning, reflected in their course grades and scores, and the students' attitudes about the course and their responses to the Classroom Community Scale (Rovai, 2002b), which were collected anonymously in course evaluations. I collected information about the demographic characteristics of the students in each class on student information sheets distributed on the first day of class. The first of the four semesters produced a baseline measurement on all variables of interest. During the last three semesters, I implemented the primary innovation, consisting of breaking the class into permanent "home teams" (formed by seating proximity, not randomized) that functioned primarily outside of class time. In all four semesters, the students completed a long-term group writing assignment, a few aspects of which were modified, one at a time per semester, over the three treatment semesters.

Design

The design of the present study is specifically a *nonequivalent control group quasi-experimental design.* This variant of quasi-experimental research (abbreviated hereafter to NEGD for nonequivalent group design) is purportedly the most common design in many areas of social research, particularly because of its value in evaluating similar, intact groups, such as classes (Campbell & Stanley, 1963). In this design, intact groups that have not been randomly assigned are compared, with the understanding that they are similar but not statistically equivalent. The fact that the groups are not randomly assigned does potentially mean that they could be somehow systematically different from

each other from the start (an internal validity threat called *selection bias*). There are particular threats to internal validity that this design controls for, and the few to which it is vulnerable can be accounted for by adding design features or adjusting the statistical analysis.

Participants

Undergraduates (N = 492) at a large state university in the Midwest United States participated in the present study by virtue of being enrolled in the classes included in this analysis. Students' course records from four consecutive semesters of the *Introduction to Social Psychology, PSYC 288* course were used as data for this investigation. The sample otherwise generally reflected the ethnic (typically white) and age (average about 20 years old) composition of the university. Several students were removed from the sample for having too much missing course grade data: 10 students were missing three or more exam grades out of four, as well as the writing assignment grades. Students missing all of the demographic information collected at the beginning of the semester (N = 28) were excluded from the data set because the comparability of the students across semesters was determined using a compilation of that demographic information. While some of the variables from that information could support imputation for a small number of cases with missing data (e.g., four students neglected to indicate how many credit hours they were taking for the semester, and those cases were imputed using the mean credit hour enrollment for the sample), the extent of the information missing from the 28 excluded individuals was far beyond what would be possible to impute with any realism. Minus the dropped cases, the final sample size consisted of 454 students (N = 98 Fall 2007, N = 92 Spring 2008, N = 131 Fall 2008, and N = 133 Spring 2009). Nineteen percent were

first-year students, 33% were sophomores, 27% were juniors, 19% were seniors, and 2% listed their class as “other,” which typically indicated that they were graduate students or non-degree-seeking. Out of the entire sample, 8% indicated that they were to graduate at the end of the semester. The majority (66%) of the students were not psychology majors. Students were enrolled in an average of 14.48 credit hours (*SD* = 2.27) for the semester. Out of the entire sample, 66% indicated that they also held a job outside of school (*M* = 14.63, *SD* = 13.15 hours worked per week).

Some of the hypotheses were tested using the anonymous data from the course evaluations. While the same students fill out the course evaluations as those used in the analyses of the course grades, the course evaluation sample is a subset of the overall study sample because some students did not complete course evaluations. Course evaluations, administered at the end of the semester, are not compulsory, and compliance depends on students’ attendance the day the evaluations are handed out. Therefore, I will describe the qualities of the sample of course evaluation participants, separately, here.

Out of the full study sample, 340 students completed course evaluations. Four cases were dropped because the students filled out the evaluation form incorrectly, reporting impossible responses to more than half of the questions. (Impossible responses either indicate that they bubbled in a letter option that was inappropriate for the question, like answering with a 5 for a yes or no question coded 1 or 2, or that the automatic reading of the scantron form produced an error due to an improperly bubbled response.) For one student who reported an impossible response to only three categorical questions (each a binary variable: psychology major or not, recommend the course or not, and recommend the instructor or not) those responses were treated as missing data and

imputed using the sample mode for those questions. Five cases were dropped because the students answered fewer than half of the questions. Aside from those nine total dropped cases, there were very few and widely scattered missing data points. Only those missing from the demographic variables used as covariates for the hypothesis tests were imputed, each using the mode for those variables. Imputation was required for fewer than five cases within each of the covariates used (i.e., students' year and major). For other questions, the occasional impossible response was merely deleted, switching it to a system-missing value for that question or variable only. Consequently, some of the analyses reported below will have different degrees of freedom because cases with missing values for variables used in the analysis will have been dropped. Of the 331 valid cases (N = 71 Fall 2007, N = 68 Spring 2008, N = 92 Fall 2008, and N = 102 Spring 2009), 18% were first-year students, 34% were sophomores, 28% were juniors, 17% were seniors, and 3% listed their class as "other." The majority (63%) of the students were not psychology majors. As for students' expected grades in the class, 84% predicted receiving an A or a B for the course, which is an interesting contrast to the actual 55% of the students who had earned an A or a B according to the actual course records, considering they could view their grades online throughout the semester.

Materials and Procedure

All four classes used the same lecture material, in-class activities, exams, and assignments. Policies, point values and breakdowns, and organization were constant across semesters as well. All course material was developed in previous semesters of my teaching this class, so it had been vetted before the terms covered in this investigation.

The first of the four semesters was used as the baseline measurement for students' performance and attitudes concerning the course curriculum. The following three semesters included the "home team" component as an intervention designed to improve students' sense of connection to each other and to facilitate learning of the course material. The team innovation was incorporated as an ancillary course component, wherein students' interactions took place almost entirely outside of class and students' grades were only tangentially related to their team participation. The last of the three treatment semesters included an additional three occasions where the teams briefly interacted as such during class time. This fourth semester addition was included in order to make the teams more salient by specific, face-to-face interaction within the classroom.

Home Team Innovation

The teams were designed in accordance with implicit social psychological recommendations for group dynamics contributing to learning. The following group characteristics were purposely chosen as conducive to my teaching objectives, and are numbered to more clearly delineate the separate points. The following list is compiled from several components used in published studies incorporating teams or groups into their classes, and includes several other components that make this strategy unique. An asterisk denotes the components that have not been included in previous research related to this teaching method, and are therefore unique to my innovation.

1. *I made the students' course grades independent from their team activities in order to avoid producing feelings of resentment for perceived loss of control over their own grade. Both Barfield (2003) and King and Behnke (2005) report one of students' specific, consistent complaints with group work is that group grades or

grades based on group work wrest control over one's grade out of the students' own hands. Because my idea is to use the social support benefits of group work to simultaneously support learning, I wanted to alleviate this concern for students from the outset.

2. *I incentivized team interaction using positive reinforcement (i.e., extra credit) to motivate team interactions (instead of punishing those who refrained), in small amounts to avoid the over-justification effect and any threat of coercion. The research by social psychologists and others investigating the effect of reward on motivation helped inform this decision (Festinger & Carlsmith, 1959; Lepper, Greene, & Nisbett, 1973; Rand, Dreber, Ellingsen, Fudenberg, & Nowak, 2009).
3. I divided the class into permanent teams to allow time for long-term, stable groups to develop across a meaningful span of time (i.e., the semester), in order to solidify group identity formation. There were typically 10-15 students in each team. Summers, Beretvas, Svinicki, and Gorin (2005) note that permanent teams help students develop a stronger sense of classroom community, and multiple other researchers use permanent groups across the semester to facilitate cooperative learning (Lancy & Rhees, 1994; Pimmel, 2003; Wolfe, Lee, Wu, & Gould, 2003).
4. I created the teams based on where students sat in class (producing de facto self-selected groups). I waited until the second week of class to form the teams in order for them to settle where they wished and to avoid late drop/adds. From the first day of class, they knew when and how I would form the teams, and I reminded them each day of the pending formation up to the day I did it. Several

researchers use the second week of class to form permanent teams, in order to maximize the students' stability in the class (both in terms of them "settling in" and in terms of enrollment shifts (Lancy & Rhees, 1994; Wolfe, Lee, Wu, & Gould, 2003). McKinney and Graham-Buxton (1993) formed teams by where students sat in the room. All of these researchers also used self-chosen teams in order to achieve heterogeneity in terms of demographics, ability, experience, and personal characteristics.

5. I had the students create unique team identities by choosing names, mottos, and mascots. I took a digital picture of each team and posted it on the front webpage of each team's section of BlackBoard so they could see all of their teammates' faces when communicating online. The researchers who incorporated these tools found that students appreciated the chance to distinguish themselves as unique collectives (Barfield, 2003; Johnson & Johnson, 1994; Lancy & Rhees, 1994). Johnson and Johnson contend that such organizational accessories improve the sense of interdependence between students, which subsequently improves their performance and achievement in the class, and social psychologists would agree that identity is important to the development of entitativity.
6. I provided and encouraged multiple modes of communication between team members in and out of class time, both in person and online through BlackBoard, synchronously and asynchronously. Synchronous online communication is between people who are simultaneously online and interacting in real time (e.g., chat or instant message). Asynchronous communication is when one person responds to another person at different times and the message remains visible for

later reference (e.g., email or discussion boards). The online communication methods included a team email list, discussion board, file sharing capabilities, and chat room (all restricted access, limited to team members and me). Their online team location showed their name, motto, and the closest I could find to a picture of their chosen mascot, and their discussion boards showed the team picture on the front page. Rovai (2002a, 2007, for two examples) has an extensive body of research on the benefit of multiple modes of synchronous and asynchronous communication on building community. His studies note that the relationship between structure and classroom community can be negative – the more structured the interaction, the more classroom community development can be stifled. Therefore, I encouraged much of the communication between the team members to be extemporaneous and on their own time. Rovai also insists that communication should be allowed to be both task-based and socializing-oriented, separately, to engender feelings of connectedness among working groups.

7. *I provided salient out-groups via competition between teams by rewarding the highest performing team, based on the team's averaged exam grade, with one point extra credit per student per exam. This practice also produced a superordinate goal for team members motivated to earn the bonus point by doing well on their exams. I emphasized the competitive aspect by showing the class a graph with all the teams' averaged scores after each exam. Research by Locke and Latham (2002) and Sherif (1956) on group dynamics and the power of superordinate goals, out-group competition, and feedback in terms of team success informed this decision. Johnson and Johnson (1994) suggested setting

achievement levels and rewarding teams that meet the criteria, and I revised their suggestion to incorporate the element of competition by only rewarding the team that earned the highest score. Mullen and Copper (1994) propose that the practical way to increase feelings of entitativity in a group is to work directly on increasing social integration, and they suggest that instructors either create programs that draw the students together to build relationships, or elevate the students' pride in their group using competition and social comparison.

8. * Independent of overall team exam outcome, I rewarded team exam studying efforts by awarding extra credit (up to two points per exam) to individual team members, based on their contribution. The in-person team studying effort was self-reported, and the online team studying efforts were via discussion board, file share, and chat room, of which there was a permanent record on BlackBoard. After each exam, I displayed class-wide statistics showing that the students who contributed to their team studying effort earned higher exam scores (which they always did) compared to those students who did not contribute. Steinbrink and Jones (1993) incorporated cooperative test review into their study, but they used formalized review activities and structured assignments to prepare for their exams. Furthermore, they used group rewards and collected individual grades on review assignments in order to reduce loafing. I modified their approach to reward individual contributions to team efforts in order to capitalize on students' self-interest (i.e., earning points for themselves) while benefiting the group. This element also adapted the procedure in the study by Rand, Dreber, Ellingsen, Fudenberg, and Nowak (2009), who documented the long-term benefit of

rewarding contributions to the common good for groups of people who were working together over time.

9. I told students who missed class that they needed to get class notes from a teammate, because the TA and I did not provide notes or lecture slides. This gave students another, non-test-related reason to turn to their team for resources. Lancy and Rhees (1994) required this same dependence in their teams.

Student Information Sheets

On the first day of the semester, all students completed a brief information sheet that requested basic demographic information (Appendix A). Students provided their name, year in school, whether they were expecting to graduate at the end of the semester or not, their major, and their contact information. Students also reported whether or not they had a job outside of school and how many hours a week they typically worked, the number of credit hours in which they were currently enrolled, the psychology courses they had already completed, and the reason they had enrolled in the course. Students were also given a place to write what typical elements of classes they did or did not like (e.g., lectures, group activities, discussion, movies, etc.), and an interesting fact about themselves.

Overall Learning Outcomes

For my Social Psychology classes, my stated learning goals were threefold: students should (a) understand the major theories and principles in social psychology, (b) learn about how social psychologists use empirical research to test their ideas, and (c) be aware of how social psychology can help them understand themselves and the world around them. In addition, my instructional objectives required that students think

critically and analytically, write at a college level, recognize and evaluate assumptions and controversies within the discipline, and appreciate diversity in opinion and background. The group writing assignment described below, and the way that I parsed its grading, allowed me to assess my instructional objectives. Furthermore, the writing assignment ultimately aids students' understanding of the principles, theories, and applications that are assessed by my exams. The learning goals and instructional objectives were complementary –students' ability to perform the skills described in my instructional objectives would interact with their ability to comprehend social psychological theories, principles, and research methods, and apply them to "real world" or personal examples. Exams in my class included factual, conceptual, and application level questions over the material covered in class lectures and activities and assigned readings. Each chapter's section on the exam was worth roughly equal points, each unit exam covered three or four chapters, and there were four unit exams across the semester. The questions remained constant across semester, and I collected the exam packets to prevent questions from being available to future students before taking the exams themselves. Exam grades therefore represented students' learning outcome for the topical content of the class.

The final course grade included students' exam scores, their writing assignment scores, and their participation and attendance in class. The final course grade was designed to represent how well the student met my content learning goals and instructional objectives for the course. Therefore, the class grade served as another indicator of their learning outcome in addition to the exams.

Semester-Long Group Writing Assignment

All four semesters included a semester-long small group writing project in the form of a movie content analysis, consisting of individual students' essays being discussed online. Students were asked to use analytical thinking to identify psychology concepts portrayed in the movie (each group chose a movie from a list I provided, all popular, feature-length, fictional films; see Appendix B), think critically about the psychological topic's depiction in the movie compared to what psychological research says, and write an essay explaining the psychology construct and how it was applied in the movie example. Students' discussion of the essay was carried out in informal, conversational writing online in discussion board format, but to receive full credit for each post, the writing had to reflect the standard of content detail and writing quality that the essays were held to. In the course of the discussion, students confronted any controversial or ambiguous elements relevant to their topics and dealt with other students' opinions when discussing the topic essay over the course of each week. Students spent the last of the seven weeks of the assignment discussing overview/review questions about their experience with the assignment in general, online as well.

The purpose of the writing assignment was threefold. First, the students were supposed to learn how to identify principles of social psychology as they are encountered outside of class material, and evaluate the principles' depiction in popular media compared to research and theory. Second, the students were supposed to learn how to discuss the principles with classmates in a way that demonstrated their facility with the material, ability to express what they knew about the topic, and to expand upon or correct what other students said about the topic. This process ultimately requires the students to have a deeper understanding of the psychology topics they discuss, how various social

psychology topics are connected, and how to apply the concepts to themselves and the world. Third, the students were supposed to learn how to work with their group members to fulfill the assignment, both by regulating their own involvement to earn their own grade, but also by dealing with other students' opinions, work habits, and perspectives.

The small groups for the writing assignment were formed based on where students were sitting in class. In the semesters with teams, this meant that the small groups for the assignment were split out of the teams. In each of the semesters, one element of the group writing assignment was changed based on the feedback about the assignment from students, in order to improve the assignment's ability to meet my instructional goals. In the second semester, the student essays were posted incrementally (1 per week) to manually spread out discussion evenly across the length of the project, instead of all essays being posted initially and discussion ranging across all essays simultaneously, as in the first iteration of the assignment. In the third semester, a peer-review element was added to the grading scheme of the assignment, in which a small portion of the overall project grade was determined by averaging each discussion group member's assessment of each other's discussion quality. In the fourth semester, a graded sample discussion, showing the actual calculation used to determine the discussion grade, was posted as a model at the beginning of the assignment. Each time an element was changed, that change persisted across subsequent semesters. Each semester, students received scores on the initial essay and the discussion listed in the gradebook separately, even though both components were technically part of the same assignment.

The writing assignment was graded in two sections. Students received one portion of their score based on the written essay about the psychology topic of their

choosing. This score included both a consideration of the content of the paper as well as the basic writing mechanics expected in college-level writing, and was worth 15 of the 60 points possible. The larger portion of the assignment score was for the online discussion. Students' individual contributions to the group discussion were evaluated based on the quality of each of their comments (e.g., whether it clarified or corrected information, contributed new information about the topic, made a new connection with another topic in psychology, or relayed a personal experience with a real-world application of the principle under consideration). The accuracy, length, detail, and clarity of each student's comments were factored in the rating each comment received, and the cumulative ratings of the week's worth of contributions made up the weekly discussion score. The weekly discussion scores were compiled into one score for the discussion portion of the assignment. Students were instructed to check in on the discussion board regularly and ask questions to move the discussion forward (as opposed to everyone agreeing with a sentiment and stopping there). Students were also told that, if there were less communicative members of their group, they could protect their own discussion grade by posting a "status check," indicating they had checked the board to see if there were new responses to a question they asked and, if there were none after a reasonable delay (e.g., a full day with no response), the student could still earn credit for conscientious participation. This strategy was designed to give students more control over their own grade on the assignment even if there were people in their group who were not contributing equally. All comments were rated on a scale from *0: no response or off-topic* to *5: extensive participation*, and a detailed grading rubric with a description of each level of participation quality was provided at the beginning of the assignment.

Students in all semesters were advised that waiting until the metaphorical last minute to join the discussion had a concrete and adverse effect on their grade. They were reminded frequently that they were to be discussing over the course of each week (or all the weeks delineated for discussion in the first semester of the assignment), and that their participation grade would be lowered if they did not contribute until the last two days of the week. This policy was included to counteract students' tendency to procrastinate and then attempt to fit a week's worth of asynchronous discussion with other people into an impractically short amount of time.

The discussion portion of the grade was split when peer review was included in the grading scheme during the last two semesters. Students rated their own and all group members' performance twice over the course of the discussion portion of the assignment. Students used the same rating scale the teaching assistant and I used to grade their discussions, considered three weeks at a time (there were six total weeks of discussion excluding the week for review questions), and wrote everyone's scores on a sheet they each turned in anonymously. The teaching assistant and I averaged the ratings provided by everyone in a group for each person in the group and that score was included in each student's overall discussion grade. The average was used to mitigate the effect of students who either failed to turn in their peer review form or who rated themselves and/or everyone with the highest or lowest possible score. Ten out of the 45 points for the discussion portion of the assignment were from the peer review, meaning that the peer review comprised 16% of the entire assignment grade.

Rovai's Classroom Community Scale

Alfred Rovai's *Classroom Community Scale* (CCS; 2002b; See the questions at the end of Appendix C) was used with permission to evaluate the students' sense of community at the end of each semester. If the team intervention was effective in producing the greater sense of connection between students, then the results of that measure should demonstrate it. Rovai's CCS was chosen because it is a widely used measure of classroom community, and it affords both an omnibus measure and two subscales ("connectedness" and "learning goals"). It is furthermore appropriate because it has been used for measuring online community too, for distance students. In my classes, apart from a few brief occasions face to face in class, team members' primary mode of communication was online. For the students' collaborative exam studying, the main venue was the BlackBoard environment. In addition to the team activities, the group writing assignment was an online assignment completed outside of class time. Therefore, Rovai's scale is appropriate for both environments. The CCS has a Cronbach's alpha value of 0.93, and an equal-length split-half coefficient of 0.91 for the omnibus scale. The connectedness subscale has an alpha *and* an equal-length split-half coefficient of 0.92, and the learning subscale has an alpha of 0.87 and an equal-length split-half coefficient of 0.80. The questionnaire is comprised of 20 questions answered on a 5-point Likert-type scale from *A = Strongly Disagree* to *E = Strongly Agree*, with *C = Neutral or No Opinion* as the midpoint. The response options were denoted by letters instead of numbers in order to facilitate the use of a bubble sheet to record participants' answers. The scale produces both an omnibus score and two subscale scores by summing the ratings of the questions belonging to each set. For this setting, the letters were converted to numbers (e.g., A = 0, E = 4, etc.), with 10 negative items reverse scored.

The range for the omnibus scale is 0-80, and the subscales ranged from 0-40, with higher numbers indicating a stronger sense of community.

Course Evaluations

At the end of each semester, students filled out course evaluations anonymously. These questionnaires included questions about the class overall, instructor specifically, course materials including the textbook, BlackBoard, and assignments (Appendix C). The majority of the questions on the questionnaire were rated on numeric scales and students used bubble sheets to record their answers. A sample question was, "The exams appropriately covered readings and classroom presentations," to which students replied by indicating their degree of agreement with statement on a 7-point Likert-type scale from *A = Strongly Disagree* to *G = Strongly Agree*, with *D = No opinion or Neutral* as the midpoint. Response options were labeled with letters instead of numbers in order to correspond to the bubble sheet format, and to avoid implying a potentially influential numeric value for students' opinions. Other questions asked students to compare the course and instructor to similar others in their experience, report their prediction of their final course grade, and indicate their year in school and whether or not they were a psychology major. Students also reported whether or not they would recommend the course or the instructor to a friend. Students also had an opportunity to answer free-response questions about what they liked and disliked about the course and individual course components, as well as the instructor's teaching style in particular. The course evaluation form had 30 questions with categorical response options, and six free-response questions.

At the end of the course evaluation form, students answered the aforementioned Classroom Community Scale (Rovai, 2002b), using the same bubble sheet to record their answers. One of the subscales of the CCS is "Learning Goals," which asks questions that are thematically similar to questions on the course evaluation, but the subscale questions are worded to emphasize more interpersonal elements of learning. The course evaluation questions ask more about how well the instructor communicated or how helpful the course components were, whereas the CCS subscale questions ask about students' sense of their learning in the context of other students' presence. A sample question on the CCS, which is stated negatively and therefore reverse scored, is "I feel uneasy exposing gaps in my understanding."

Informing Students of Use of Records

Research using data collected in the normal course of conducting a class is exempt in Institutional Review Board (IRB) reviews. The IRB classifies this evaluation as a secondary data analysis of archival information. Because students participated in the normal course of completing the class, I incorporated procedures based on existing protocols for peer review and classroom research to inform them of the potential uses of their information. First, there was a statement in the syllabus describing what records I might collect from class and how it could be used in future research. Second, at the end of the semester, I gave my students a sheet repeating the info from the syllabus. I made clear that future use of their class records had nothing to do with their course grade in any way; their data would be aggregated with everyone's, anonymized, and there would be no benefit for including their records or penalty for excluding their records in future research. The last protection for students was that once the data files were aggregated,

the students' names were removed from the data sets. Data from course evaluations were already anonymous.

Chapter 3: Analysis

The data used in this analysis were collected in the normal course of teaching the class. Some of the data originated from known students, such as the grades and student information sheets, and other data, including the responses to the CCS, were collected anonymously on the course evaluations. I will describe the form of the data used in these analyses here, and then organize the results by the hypotheses they are testing below. All analyses were done using IBM's Statistical Package for the Social Sciences (SPSS), version 18.

Student information sheets provided categorical information about named students' year in school (i.e., first-year, sophomore, junior, senior, and other or unspecified), major (coded as binary *psychology major* or *non-psychology major*), whether or not they were graduating at the end of the semester (binary *yes* or *no*), the number of credit hours enrolled in during the semester (a continuous variable), whether or not they had employment outside of school (binary *yes* or *no*), how many hours they worked on average per week in their outside job (a continuous variable, with zeros entered for those indicating they did not have an outside job), and whether or not they mentioned "group work" in their answers to the questions of what common elements of courses they liked and disliked.

Course grades included the scores students received on each of the four exams and the writing assignments as continuous variables. Final course grades were expressed as a total number of points earned out of the amount possible in the class, and are

continuous as well, though that metric could also be expressed as letter grades, which are categorical. The overall online group writing assignment grade was points-based as well, and the separate scores for the essay and the discussion were also included in the data set for analyses looking at only one component. In the last two of the treatment semesters, the discussion participation component also included the points from the peer review component. Students earned bonus points, available up to the same limit for each treatment semester, by engaging in activities with their home teams. These extra credit points were included in the overall course grade calculation out of zero possible (i.e., with a denominator of zero), but as a category of data, were continuous like other course points.

On the anonymous course evaluations, students rated how good they thought the course was and how helpful they found class activities on a continuous scale. Students provided categorical answers about their year in school, major, anticipated grade, and whether or not they recommend the course to other students. Finally, students completed the Classroom Community Scale (CCS; Rovai, 2002b), which produces continuous scores on both the omnibus measure and both subscales.

Initial Group Similarity

I analyzed the similarity of the comparison groups (i.e., semesters) because of my use of non-equivalent groups (i.e., groups that are not randomly assigned) for this study. Because statistical equivalence, which is only produced by true random assignment, was impossible, initial group similarity was determined using the student information sheet data. This provided demographic information (categorical year in school and binary psychology major or not), enrollment level in credit hours, stated dislike for group work

(binary), and extracurricular work responsibilities in average hours per week, for a total of five variables functioning like covariates. Covariates are factors that are unrelated to the independent variable (IV), but are suspected to be related to the dependent variable (DV); controlling for them increases the power of the analysis of the primary relationship between IV and DV. Initial imbalance across semesters on binary variables was tested using a logistic regression with the binary covariate as the outcome and the semesters dummy coded as a categorical predictor with the first semester as the baseline (Zanutto, Lu, & Hornik, 2005). A nonsignificant result indicates adequate equivalence across the semesters, meaning that the semester group membership cannot be used to predict the covariate outcome. The binary covariates, psychology major and stated dislike for groups, were adequately equivalent across semesters, as indicated by the nonsignificant and poorly fitting model: the model testing balance for majors produced a highly nonsignificant fit, *–2LL* = 578.91, *Cox and Snell* R^2 = 0.009, model Chi-square statistic $\chi^2(3, N = 454) = 3.99$, $p = .262$, and the model for group dislike was similarly ill-fitting, *–2LL* = 627.05, *Cox and Snell* R^2 = 0.004, model Chi-square statistic $\chi^2(3, N = 454) = 1.89$, $p = .595$. See Table 3.1 for odds ratios and Table 3.2 for descriptive statistics on the covariates.

Table 3.1

Logistic Regression Values for Binary Covariates across Semester

Semester (Predictor)	Psychology Major or Not *Odds Ratio* (95% CI)	Stated Dislike for Group Work *Odds Ratio* (95% CI)
Fall 2007		
Spring 2008	*1.051* (.577-1.913)	*1.395* (.788-2.471)
Fall 2008	*.775* (.440-1.365)	*1.339* (.791-2.266)
Spring 2009	*1.305* (.757-2.248)	*1.117* (.661-1.887)

Logistic regression results reporting Exp(B) and 95% CI for both of the binary covariates, with semesters dummy coded where Fall 2007 is 0. Odds ratios indicate the effect size of the association between two binary variables.
Note: All values nonsignificant. CI stands for Confidence Interval.

The initial balance for continuous covariates was tested using a one-way ANOVA, with semester as the independent variable and the covariate as the dependent variable (Zanutto et al., 2005). Again, non-significant differences indicate that the semesters are adequately similar on both continuous covariates (i.e., hours enrolled and hours of extra-curricular work outside of school). Neither the number of hours students were enrolled in for the semester, $F(3,450) = 1.236$, $p = .296$, $\eta^2 = .008$, nor the number of hours students worked outside of school, $F(3,450) = 0.909$, $p = .437$, $\eta^2 = .006$, produced significant differences across the semesters.

Table 3.2

Original Descriptive Statistics on the Variables that Comprised the Propensity Score Covariates

	Semesters				
Variables	Fall 2007 (N = 98)	Spring 2008 (N = 92)	Fall 2008 (N = 131)	Spring 2009 (N = 133)	Total Across All Semesters
Psychology majors	33 (34%)	32 (35%)	37 (28%)	53 (40%)	155 (34%)
Mean hours enrolled (SD)	14.87 (1.82)	14.42 (2.37)	14.37 (2.23)	14.34 (2.51)	14.48 (2.27)
Mean hours worked outside school (SD)	16.35 (13.50)	15.01 (11.75)	13.64 (12.45)	14.08 (14.42)	14.63 (13.15)
Stated dislike for groups	43 (44%)	48 (52%)	67 (51%)	62 (47%)	220 (49%)
Year in School					
First-year	10 (10%)	25 (27%)	10 (8%)	41 (31%)	86 (19%)
Sophomore	33 (34%)	31 (34%)	45 (34%)	40 (30%)	149 (33%)
Junior	36 (37%)	20 (22%)	35 (27%)	32 (24%)	123 (27%)
Senior	16 (16%)	14 (15%)	38 (29%)	18 (14%)	86 (19%)
Other	3 (3%)	2 (2%)	3 (2%)	2 (2%)	10 (2%)

The only covariate that was not functionally equivalent across semesters was students' year in school, the initial balance of which was evaluated using Pearson's chi-square test of independence. Analysis revealed a significant relationship between students' year and their semester group, $\chi^2(12, N = 454) = 41.955$, $p < .001$, *Cramer's V* = .176, $p < .001$. Cramer's V indicates the effect size, and it means that the relationship is relatively weak (the determination of its strength is similar to correlation coefficients) but still highly significant. Another way of describing the relationship is using Goodman and Kruskal's lambda, which measures how much error is reduced when one group membership variable is used to predict group membership in the other, with a perfect

relationship indicated by a value of one. When the semester membership is the dependent variable, students' year does significantly predict group membership, even if weakly λ = .093, p =.04. Because this covariate has an initial imbalance across semesters, group equivalence will be adjusted in the hypothesis tests described below, using propensity scores (described in detail below) as a scalar representative of students' semester membership. This allowed the hypothesis tests to proceed while reducing the loss of internal validity due to initial group dissimilarity resulting from the quasi-experimental design.

Using extraneous variables as predictors of participants' treatment group membership (treated as the outcome variable) allows the researcher to estimate the probability of each participant being in a condition (Yanovitzky, Hornik, & Zanutto, 2008). This procedure creates a *propensity score* for all cases (see Rosenbaum & Rubin, 1983, for an in-depth theoretical explanation of the procedure), and can then be used to control all subsequent analyses of the treatment effect. The propensity score operates as the individual covariates would, but combined into one variable to preserve degrees of freedom and create different treatment groups that are similar on the observed covariates. In randomly assigned groups, one assumes that the covariate distributions are equally probable in each condition, and thus can be compared across groups. Propensity scores in a non-randomized study provide a value that can be used to compare participants on a standard scale based on their covariate values – a person from the control condition is compared to a person in the treatment condition with a similar propensity score. A propensity score essentially adjusts analysis of the treatment effect to control for any potential selection bias, by comparing participants that are similar in covariate qualities

but differ in their treatment exposure (Yanovitzky et al., 2008). In the present study, propensity scores were constructed using the five covariates described above.

The method used here to produce the participants' propensity scores differed slightly from more common applications of this method, due to the present study having four comparison groups instead of a binary comparison (i.e., treatment versus control conditions). While propensity scores have been in widening use for the past 30 years (since established by Rosenbaum & Rubin, 1983), usage with multiple comparison groups is more recent and therefore less established (Yanovitzky, Hornik, & Zanutto, 2008). However, researchers are doing more with multiple group studies (see Imbens, 2000; Zanutto, Lu, & Hornik, 2005). Binary group propensity scores were estimated using logistic regression, where the five relevant demographic covariates serve as the predictors, and participants' belonging or not belonging in each semester was the outcome. The covariates included as the predictors of the semester membership represent the qualities expected to be relevant to students' performance in the class, but are independent of the semester in which they participated. Belonging in a semester was dummy coded, with 1 given to participants enrolled in the semester being analyzed, and 0 given to all others. Each semester served as the outcome for its own regression. SPSS saves the predicted probability of belonging in the outcome semester in the dataset as a byproduct of the analysis, and this value is what I used as the propensity score. Therefore, I had four propensity scores for each participant (each representing the likelihood of that person being in that semester). When estimating propensity scores, the model is built according to the *hypothetical* relevance of the *pre-selected* covariates and any of their logical interactions (Yanovitzky et al., 2008) - in fact, Zanutto and her

colleagues (2005) explain that even non-significant predictors should not be removed from the model for this reason.

Once the four separate propensity scores were estimated, the propensity score distribution for participants in the semester under investigation was compared to the distribution of the scores for participants who were not under investigation (i.e., all other participants). For example, the distribution of propensity scores (i.e., the probability of being enrolled in Fall 2007) for all the people who were enrolled in Fall 2007 was compared to the distribution of propensity scores (still the probability of being enrolled in Fall 2007) for all of the other participants enrolled in other semesters. Based on this comparison, any participants whose scores did not overlap with the other group (e.g., people from Fall 2007 whose scores did not have matching scores among the other students in the sample) were culled from the dataset (Zanutto, Lu, & Hornik, 2005). Each semester underwent this comparison, in chronological order, without replacing previously culled participants for subsequent comparisons. At the end of this matching process, 11 participants had been cut from the sample (Fall 2007 $N = 0$, Spring 2008 $N = 6$, Fall 2008 $N = 1$, Spring 2009 $N = 4$) for not having similar enough propensity scores to compare across groups. This left the study sample with 443 participants (Fall 2007 $N = 96$, Spring 2008 $N = 91$, Fall 2008 $N = 124$, Spring 2009 $N = 132$).

The next step in balancing the sample using the propensity score is to stratify the samples into equal quintiles (Yanovitzky, Hornik, & Zanutto, 2008). While there are several potential strategies for balancing the sample's propensity scores (Austin, 2011; Hahs-Vaughn & Onwuegbuzie, 2006), the stratification method is preferable because it allows the researcher to retain the bulk of the participants, as opposed to, for example,

one-to-one matching where any participant without an equivalent participant in the control group would get cut from the sample. Because the enrollment in my four semesters was slightly larger in my last two semesters, I wanted to keep as many of my participants as possible. Each of the four semesters' propensity scores were sorted into ascending order, then separated into equivalent sized groups of participants using SPSS's visual binning function. This means that at the lowest end of the propensity score distribution, the fifth of the sample with the lowest scores were grouped together as similar, and then the next fifth of all participants with the next higher propensity scores were sub-classified, and so on for the whole sample. See Table 3.3 for the propensity scores by quintile per semester summarization to help visualize the stratification process.

Table 3.3

Propensity Scores by Quintile per Semester

	Semester			
Quintile	Fall 2007	Spring 2008	Fall 2008	Spring 2009
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
1	.066 (.042)	.130 (.027)	.104 (.039)	.187 (.020)
	N = 88	*N* = 89	*N* = 88	*N* = 88
2	.158 (.016)	.172 (.007)	.209 (.025)	.229 (.013)
	N = 90	*N* = 88	*N* = 90	*N* = 90
3	.211 (.013)	.195 (.006)	.286 (.018)	.259 (.008)
	N = 86	*N* = 88	*N* = 89	*N* = 88
4	.265 (.020)	.224 (.012)	.362 (.027)	.325 (.038)
	N = 90	*N* = 87	*N* = 87	*N* = 86
5	.371 (.055)	.287 (.027)	.472 (.060)	.473 (.045)
	N = 89	*N* = 91	*N* = 89	*N* = 91

Note: The values listed for the propensity score means denote probability, and therefore are out of a maximum possible value of 1.

The final step in balancing the sample using propensity scores is rechecking the covariates' distribution across semesters (Yanovitzky et al., 2005). The previously described methods were employed with one additional step. While the initial test had the covariate as the outcome (or DV) and semester as the predictor (or IV), here the quintiles were included as predictors as well. Any model with a significant main effect for semester, or a significant interaction with the quintiles and semester, indicates that the covariate is not balanced across semesters. These final checks revealed no significant imbalance for the major or non major, hours enrolled, hours worked outside school, and stated dislike of group work. Students' year in school, which was the original concern, also revealed no significant differences across semester. This test was performed using ordinal logistic regression with students' year as the outcome, and semester and all four propensity score quintiles as the predictors, including all two-way interactions with semester. Results indicated that the model was not a good fit for the data, $-2LL$ = 816.072, *Cox and Snell* R^2= 0.551, model Chi-square statistic $\chi^2(866, N = 443)$ = 675.244, p = 1.00, but more importantly, semester and its interactions were not significant predictors, indicating that the covariate is now balanced across semesters as if it had been randomly assigned. See Table 3.4 for the slopes and odds ratios of the ordinal logistic regression on year using semester and the four stratified propensity score variables and Table 3.5 for newly calculated descriptive statistics for the five covariates after the balancing process.

Table 3.4

Results of Ordinal Logistic Regression of Semester and Quintiles on Students' Year in School

	95% CI for Odds Ratio		
	Lower	Odds Ratio	Upper
Fall 2007	0.74	1.01	1.38
Spring 2008	0.70	0.98	1.37
Fall 2008	0.72	0.97	1.33
Spring 2009		1.00	
PropensityScoreQuintileFall2007 with Freshman	0.40	1.61	6.46
PropensityScoreQuintileFall2007 with Sophomores	0.50	1.43	4.13
PropensityScoreQuintileFall2007 with Juniors	0.61	1.27	2.68
PropensityScoreQuintileFall2007 with Seniors	0.70	1.11	1.75
PropensityScoreQuintileFall2007 with students listed as 'Other'		1.00	
PropensityScoreQuintileSpring2008 with Freshman	0.12	4.20	148.68
PropensityScoreQuintileSpring2008 with Sophomores	0.17	3.18	57.85
PropensityScoreQuintileSpring2008 with Juniors	0.26	2.39	22.02
PropensityScoreQuintileSpring2008 with Seniors	0.36	1.89	9.75
PropensityScoreQuintileSpring2008 with students listed as 'Other'		1.00	
PropensityScoreQuintileFall2008 with Freshman	0.44	1.38	4.39

PropensityScoreQuintileFall2008 with Sophomores	0.46	1.45	4.55
PropensityScoreQuintileFall2008 with Juniors	0.62	1.18	2.24
PropensityScoreQuintileFall2008 with Seniors	0.55	1.40	3.59
PropensityScoreQuintileFall2008 with students listed as ‘Other’		1.00	
PropensityScoreQuintileSpring2009 with Freshman	0.14	3.56	88.50
PropensityScoreQuintileSpring2009 with Sophomores	0.20	2.76	37.14
PropensityScoreQuintileSpring2009 with Juniors	0.34	1.86	10.19
PropensityScoreQuintileSpring2009 with Seniors	0.38	1.78	8.40
PropensityScoreQuintileSpring2009 with students listed as ‘Other’		1.00	

Note: The results of the additional 80 interaction tests that combine the elements of year, semester, and quintile are also all extremely nonsignificant, and that expanded table is available upon request.

Table 3.5

New Descriptive Statistics on the Propensity Score Covariates after Culling Mismatched Cases

	Semesters				
Variables	Fall 2007 (N = 98)	Spring 2008 (N = 92)	Fall 2008 (N = 131)	Spring 2009 (N = 133)	Total Across All Semesters
Psychology majors	31 (32%)*	31 (34%)*	31 (25%)*	53 (40%)	146 (33%)*
Mean hours enrolled (SD)	14.90 (1.82)*	14.44 (2.37)*	14.45 (1.96)*	14.34 (2.52)*	14.51 (2.20)*
Mean hours worked outside school (SD)	16.38 (13.47)*	14.84 (11.71)*	13.32 (11.93)*	13.88 (14.29)*	14.46 (12.97)*
Stated dislike for groups	41 (43%)*	47 (52%)*	62 (50%)*	61 (46%)*	211 (48%)*
Year in School					
First-year	10 (10%)	25 (28%)*	8 (7%)*	40 (30%)*	83 (19%)*
Sophomore	33 (34%)	31 (34%)	45 (36%)	40 (30%)	149 (34%)
Junior	36 (37%)	20 (22%)	35 (28%)	32 (24%)	123 (28%)
Senior	16 (17%)*	14 (15%)	34 (28%)*	18 (14%)	82 (19%)*
Other	1 (1%)*	1 (1%)*	2 (2%)*	2 (2%)	6 (1%)*

Note: Cells wherein the value changed based on the culling process are marked with an asterisk to help identity them easily.

The course evaluation dataset was appraised in the same manner as the course grades data set, to determine functional equivalence across the semesters in order to avoid selection bias. Analysis of the initial balance on two covariates (students' year and major) found a significant difference on year only. Initial imbalance across semesters on students' major (binary) was tested using a logistic regression with the binary covariate as the outcome and the semesters dummy coded as a categorical predictor, using the first semester as the baseline (Zanutto, Lu, & Hornik, 2005). A nonsignificant result indicates

adequate equivalence across the semesters, meaning that the semester group membership cannot be used to predict the covariate outcome. The students' major (psychology or not) was adequately equivalent across semesters, as indicated by the nonsignificant and poorly fitting model: the model testing balance for majors produced a highly nonsignificant fit, *–2LL* = 431.138, *Cox and Snell* R^2 = 0.011, model Chi-square statistic $\chi^2(3, N = 331) = 3.498$, $p = .321$. See Table 3.6 for odds ratios.

Table 3.6
Logistic Regression Values for Binary Major across Semester

Semester (Predictor)	Psychology Major or Not *Odds Ratio* (95%CI)
Fall 2007	
Spring 2008	1.661 (.826-3.340)
Fall 2008	1.684 (.885-3.202)
Spring 2009	1.201 (.650-2.221)

Logistic regression results reporting Exp(B) and 95%CI for both of the binary covariates, with semesters dummy coded where Fall 2007 is 0.
Note: All values nonsignificant.

The initial balance of students' year in school was evaluated using Pearson's chi-square test of independence, and analysis revealed a significant relationship between students' year and their semester group, $\chi^2(12, N = 331) = 40.078$, $p < .001$, *Cramer's V* = .201, $p < .001$. This value for Cramer's V effect size means that the relationship is relatively weak but still highly significant. Goodman and Kruskal's lambda reports that when the semester membership is the dependent variable, students' year does significantly predict group membership, though weakly $\lambda = .092$, $p = .03$. See Table 3.7 for descriptive statistics for the covariates.

Table 3.7

Original Descriptive Statistics for the Course Evaluation Covariates that Comprised the Propensity Scores

	Semesters				
Variables	Fall 2007 (N = 71)	Spring 2008 (N = 66)	Fall 2008 (N = 92)	Spring 2009 (N = 102)	Total Across All Semesters
Psychology majors	31 (44%)	21 (32%)	29 (32%)	40 (40%)	121
Year in School					
First-year	7 (10%)	14 (21%)	5 (5%)	32 (31%)	58
Sophomore	29 (41%)	21 (32%)	31 (34%)	32 (31%)	113
Junior	22 (31%)	20 (30%)	28 (30%)	24 (24%)	94
Senior	11 (16%)	6 (9%)	26 (28%)	12 (12%)	55
Other	2 (3%)	5 (8%)	2 (2%)	2 (2%)	11

Because of this inequality, propensity scores were estimated for the sample, producing four separate propensity scores (i.e., one for each semester). Each semester was evaluated for distribution overlap and 13 cases were cut from the extremity of the distributions for being without a match (Fall 2007 $N = 0$, Spring 2008 $N = 10$, Fall 2008 $N = 3$, Spring 2009 $N = 0$). This left a remaining 318 participants in the sample for the following hypothesis tests (Fall 2007 $N = 68$, Spring 2008 $N = 64$, Fall 2008 $N = 84$, Spring 2009 $N = 102$). Following the winnowing process, each of the four propensity scores was stratified into quintiles. See Table 3.8 for the propensity scores by quintile per semester summarization to help visualize the stratification process.

Table 3.8

Propensity Scores by Quintile per Semester for Course Evaluations

	Semester			
	Fall 2007	Spring 2008	Fall 2008	Spring 2009
Quintile	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
1	.079 (.024)	.133 (.000)	.147 (.066)	.239 (.000)
	N = 38	*N* = 45	*N* = 109	*N* = 67
2	.178 (.001)	.141 (.005)	.250 (.000)	.270 (.006)
	N = 90	*N* = 78	*N* = 8	*N* = 115
3	.205 (.004)	.214 (.000)	.259 (.000)	.294 (.000)
	N = 112	*N* = 28	*N* = 27	*N* = 51
4	.353 (.000)	.233 (.007)	.313 (.000)	.296 (.000)
	N = 27	*N* = 129	*N* = 67	*N* = 27
5	.353 (.000)	.289 (.045)	.355 (.038)	.552 (.016)
	N = 51	*N* = 38	*N* = 107	*N* = 58

The covariate balance was reassessed and, for both major and year, semester was not a significant predictor. Students' year in school, which again was the original source of imbalance, also revealed no significant differences across semester. This test was performed using ordinal logistic regression with students' year as the outcome, and semester and all four propensity score quintiles as the predictors, including all two-way interactions with semester. Results indicated that the model was not able to find a good fit for the data because there were different *n* in the cells and some of the combinations were not observed. This is reasonable, given that there are over 150 cells with all dependent variable levels by combinations of predictor variable values possible. Thus, SPSS reported that some parameter estimates would be dropped. However, the remaining results reported that the model's summary model fit information showed a

similar outcome to the year analysis with the previous class information dataset, *–2LL* = 181.507, *Cox and Snell* R^2 = 0.894, model Chi-square statistic $\chi^2(44, N = 318)$ = 713.747, p <.001. More importantly, semester and its available interactions were not significant predictors, indicating that the covariate was sufficiently balanced across semesters. See Table 3.9 for odds ratios of the ordinal logistic regression on year using semester and the four stratified propensity score variables** and Table 3.10 for new descriptive statistics.

Table 3.9
Results of Ordinal Logistic Regression of Semester and Quintiles on Students' Year in School

	95% CI for Odds Ratio		
	Lower	Odds Ratio	Upper
Fall 2007	0.00	1.00	659.30
Spring 2008	0.01	1.00	111.44
Fall 2008	0.00	1.00	203.85
Spring 2009		1.00	

***The results of the additional 156 interaction tests that combine the elements of year, semester, and quintile are also all extremely nonsignificant, and that expanded table is available upon request.*

Table 3.10
New Descriptive Statistics for the Propensity Score Covariates After Culling Mismatches

	Semesters				
Variables	Fall 2007 (N = 68)	Spring 2008 (N = 64)	Fall 2008 (N = 84)	Spring 2009 (N = 102)	Total Across All Semesters
Psychology majors	28 (41%)*	19 (30%)*	21 (25%)*	40 (39%)*	108
Year in School					
First-year	7 (10%)	14 (22%)*	5 (6%)*	32 (31%)	58
Sophomore	29 (43%)*	21 (33%)*	31 (37%)*	32 (31%)	113
Junior	22 (32%)*	20 (31%)*	28 (33%)*	24 (24%)	94
Senior	9 (13%)*	6 (9%)	18 (21%)*	12 (12%)	45*
Other	1 (2%)*	3 (5%)*	2 (2%)	2 (2%)	8*

Note: Cells wherein the value changed based on the culling process are marked with an asterisk to help identity them easily.

To summarize, because my design was quasi-experimental, my comparison groups were non-equivalent; in order to analyze them, I had to establish group similarity using the demographic information collected about each student at the beginning of all the semesters. Most of the covariates were already sufficiently similar across the semesters. The sole exception was students' year in school (e.g., first-years, sophomores, etc.), meaning that year in school significantly predicted which semester a student was in. Therefore, the similarity across comparison groups had to be adjusted, and I used propensity scores to balance the covariates' distributions across the semesters. After adjustment, the semesters were analyzed again to confirm that the covariates had been sufficiently controlled, and they were. From this point forward, the semesters can be analyzed as if they had had initial statistical equivalence by including the propensity score information in the hypothesis test analyses.

Chapter 4: Results

The hypotheses for the present study concern the effect of incorporating permanent home teams on students' sense of classroom community and learning of course material. The following results are organized by thematically similar sets of hypotheses.

Effect of Teams on Classroom Community

I hypothesized that students in classes including teams would report a greater "sense of community" as measured by the *Classroom Community Scale* (CCS; Rovai, 2002b), than students in the normal format class would on both subscales, "learning goals" and "connectedness," and the omnibus test. Scores on the CCS were analyzed

using an Analysis of Covariation (ANCOVA) with semester and the four semesters' propensity score quintiles as the Independent Variables (IVs). The quintiles are categorical variables representing clumps of propensity scores for each individual student. The continuous covariate was students' reported expected grade. I included students' expected grades (which tend to be blithely optimistic, as I mentioned above) in this primary analysis of the CCS because I wanted to determine if there was a sense of community for students regardless of what grade they thought they would earn in the course. Results did not significantly support the main hypothesis, although they were directionally supportive. There was not a significant effect of semester on students' sense of community, $F(3,285) = 1.089$, $p = .354$, partial $\eta^2 = .011$. Students in all three treatment semesters reported a stronger sense of community than students in the control semester (See Table 4.1 for all adjusted means), but the difference was not large enough to reach significance. Planned contrasts revealed no significant difference between the baseline control semester and the treatment semesters, a result which was corroborated by the post hoc Sidak-adjusted analysis. In the covariate analysis of the omnibus score on the CCS, students' expected grade significantly contributed to their sense of overall classroom community, $F(1,285) = 11.138$, $p < .01$, partial $\eta^2 = .03$. As students' estimate of their grade moved higher (A was coded as 1), their sense of community increased, $b = -2.552$, $p < .01$.

Table 4.1

Classroom Community Scale Scores by Semester, Including Propensity Score Adjustment

	Classroom Community Scale Scores					
	Omnibus Scale		Learning Subscale		Connection Subscale	
Semester	*M* (*SD*)	*M* (*SE*) Adj. for Covariate	*M* (*SD*)	*M* (*SE*) Adj. for Covariate	*M* (*SD*)	*M* (*SE*) Adj. for Covariate
Fall 2007	62.29	63.17	34.00	33.76	28.82	29.75
(*N* = 68)	(8.56)	(1.87)	(6.09)	(1.17)	(5.15)	(1.20)
Spring 2008	65.16	65.14	35.61	35.85	29.67	29.53
(*N* = 64)	(9.20)	(1.46)	(6.72)	(.918)	(5.89)	(.940)
Fall 2008	67.38	66.93	36.30	35.94	36.30	31.84
(*N* = 84)	(9.54)	(1.68)	(6.42)	(1.05)	(6.42)	(1.08)
Spring 2009	67.06	66.50	36.60	35.82	36.60	31.57
(*N* = 102)	(10.33)	(1.29)	(7.03)	(.811)	(7.56)	(.830)

Note: Omnibus scale ranges from 0-80, subscale scores range from 0-40, and higher values mean stronger sense of community. Covariate included in the ANCOVAs was students' expected grade in the class.

Analysis of the subscales of the Classroom Community Scale demonstrated similarly nonsignificant results. There was no effect of semester on students' sense of the classroom community helping them meet their learning goals, $F(3,285) = 1.069$, $p = .362$, partial $\eta^2 = .011$, although again, results were directionally as predicted. Neither was there a significant effect of semester on students' sense of their social connectedness goals being met by the classroom community, $F(3,285) = 1.634$, $p = .182$, partial $\eta^2 = .017$. Planned comparisons revealed no significant differences between the baseline semester and the three treatment semesters, as did the post hoc analysis (see Table 4.1). Again, students' expected grades in the class significantly contributed as a covariate to the strength of their sense of community [Learning subscale: $F(1,285) = 13.352$, $p < .001$,

partial η^2 = .045, and connectedness subscale: $F(1,285) = 6.722$, $p < .05$, partial η^2 = .023], but did not differ by semester. Students' sense of learning community and connectedness increased the higher they expected their grade to be (b = -1.754, p< .001 and b = -1.274, $p < .05$, respectively).

The results of the first hypothesis test can be analyzed without including the propensity score quintiles to illustrate the effect of the adjustment on the results. Using the data including the students that were culled in the process of balancing the covariates in the propensity score generation, I tested the difference in students' CCS scores across semesters using an ANCOVA. Students' expected grade was included as a covariate, as in the original test of this hypothesis. Results significantly supported the hypothesis. Students in all three treatment semesters reported feeling a stronger sense of community than students in the control semester, $F(3,326) = 4.216$, $p < .01$, partial η^2 =.037. Marginal means show that scores increase incrementally as the semesters progress, and both planned contrasts and Sidak-adjusted post hoc comparisons show that Fall 2008 and Spring 2009 are both significantly higher than Fall 2007 (see Table 4.2 for descriptive statistics on the three elements of the CCS). The students' expected grade also significantly contributed as a covariate to their sense of classroom community, $F(1,326)$ = 8.269, $p < .01$, partial η^2 = .025. As students' estimate of their grade moved higher, their sense of community increased, b = -2.046, p <.01. Results of an ANCOVA on differences across semester on the learning and connection subscales of the CCS also show a significant increase in scores as the semesters progress, $F(3,326) = 2.768$, $p < .05$, partial η^2 = .025 and $F(3,326) = 3.353$, $p < .05$, partial η^2 =.031, respectively. Significant differences between the control and treatment semesters are identified in Table 4.2 with

an asterisk, based on the results of planned contrasts and corroborated by Sidak-adjusted post hoc comparisons. Students' expected grade also influenced students' sense of learning and connectedness in the classroom community as a covariate. As students' estimate of their grade moved higher, their sense of the learning community significantly increased, $b = -1.659$, $p < .001$, and the increase in their sense of connection was marginally significant, $b = -.822$, $p = .069$. The difference between these results and the previously reported results of the first hypothesis test is important to point out. Because the quasi-experimental design precludes random assignment to condition, there is a risk of selection bias confounding the results. If I were to accept the unadjusted analysis results just described, I could very well be making a Type I error, because the significant difference is reflecting some systematic or pre-existing difference between my conditions. The propensity score adjustment does not change the fact that the results are still directionally supportive of the hypothesis, but the fact that those results do not achieve significance suggests that the adjustment was necessary.

Table 4.2

Classroom Community Scale Scores by Semester, Analyzed Without Propensity Score Adjustment

	Classroom Community Scale Scores					
	Omnibus Scale		Learning Subscale		Connection Subscale	
Semester	*M* (*SD*)	*M* (*SE*) Adj. for Covariate	*M* (*SD*)	*M* (*SE*) Adj. for Covariate	*M* (*SD*)	*M* (*SE*) Adj. for Covariate
Fall 2007	62.38	62.31	34.07	34.02	28.87	28.84
(*N* = 68)	(8.87)	(1.23)	(6.19)	(.781)	(5.25)	(.775)
Spring 2008	64.42†	64.47	35.26	35.30	29.26	29.28
(*N* = 64)	(10.69)	(1.27)	(7.23)	(.810)	(6.45)	(.803)
Fall 2008	67.30*	66.98*	36.34*	36.08	31.58*	31.45†
(*N* = 84)	(9.30)	(1.08)	(7.03)	(.689)	(6.31)	(.684)
Spring 2009	67.06*	67.38*	36.60*	36.85*	31.25*	31.37†
(*N* = 102)	(12.09)	(1.03)	(7.03)	(.655)	(7.56)	(.650)

Omnibus scale ranges from 0-80, subscale scores range from 0-40, and higher values mean stronger sense of community. Covariate included in the ANCOVAs was students' expected grade in the class.
Note: Semesters significantly different from the baseline semester at p<.05 are denoted with an asterisk, and p < .10 with a cross.

Effect of Teams on Learning Outcomes

The second set of hypotheses refers to the predicted impact of the teams on students' learning outcomes. Across the three semesters with teams, 68.4% of all students earned extra credit for participating in the voluntary activities for teams outside of class. First, I tested the prediction that students in classes with teams would earn higher grades on the group writing assignment than students in the semester without teams, using an ANOVA with semester as the IV. The dependent variable (DV) was the overall score students received on the group writing assignment. Also included as IVs were the propensity score quintiles for the four semesters as calculated above, derived

from students' year in school, major, credit hours enrolled, approximate hours per week of extracurricular employment, and stated dislike of group work. Their hours enrolled and hours working were included in the propensity score analysis to account for how added responsibilities outside of class might have influenced their performance on the online group writing assignment and studying for exams, which took place largely outside of class time. There was a significant effect of semester on the total assignment grade, but it was in the opposite direction than predicted, $F(3,423) = 2.899$, $p < .05$, partial $\eta^2 = .020$. Students in the baseline semester earned the highest overall assignment grade of the four semesters ($M = 46.49$ out of 60, $SD = 14.34$), and the score was lower for Spring 2008 ($M = 41.35$, $SD = 13.11$), Fall 2008 ($M = 41.62$, $SD = 12.87$), and Spring 2009 ($M = 40.10$, $SD = 15.57$). Planned contrasts between the baseline and each of the three treatment semesters were significant, and Sidak-adjusted post hoc comparisons corroborated that the only significant differences were between the treatment semesters and the baseline. There were no significant differences among the three treatment semesters.

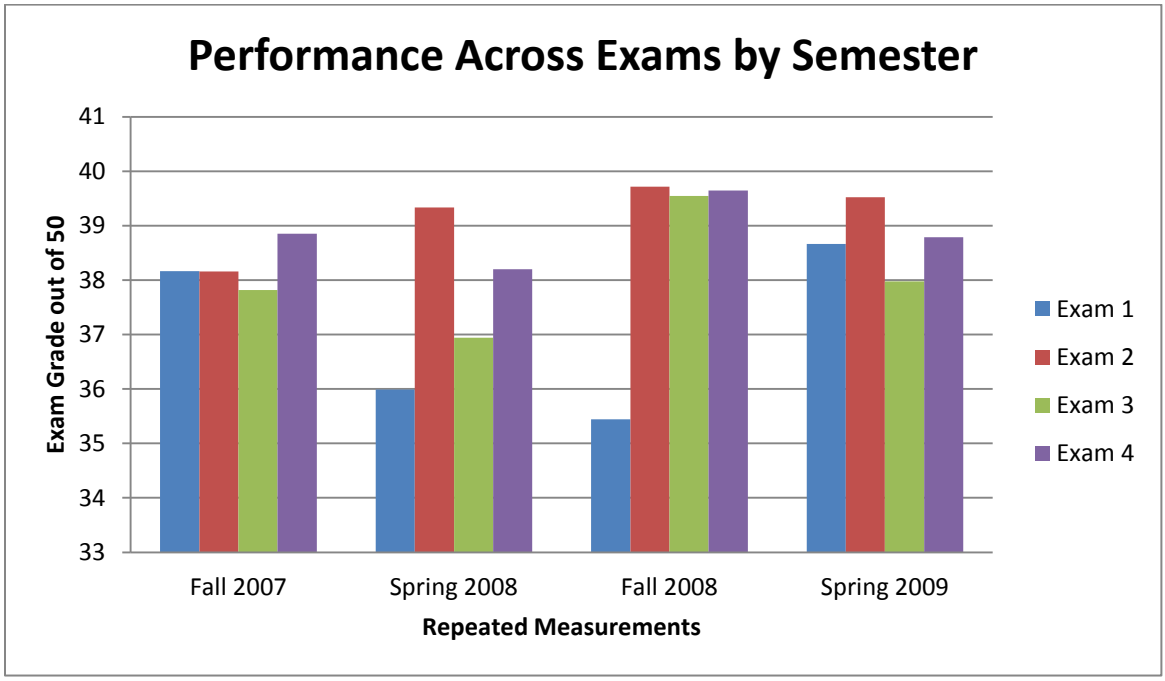
The prediction that students in semesters including teams would have higher exam scores was tested using an ANOVA with semester as the IV and students' average exam scores (e.g., average scores across all four exams) as the DV. The four semesters' propensity score quintiles were included as IVs as well. There was no significant effect of semester on students' average exam scores, $F(3,423) = 0.594$, $p = .620$, partial $\eta^2 = .004$. Students' scores on exams, out of a maximum score of 50, were remarkably stable across all four semesters (see Table 4.4). Planned contrasts and post hoc analysis concurred.

The prediction that students in semesters with teams would show more improvement across exams (e.g., over the course of the semester) than students in the baseline semester was tested using a mixed ANOVA with semester and the four propensity score quintiles as the between-groups IV. The four unit exam scores were the repeated measurement DVs. The interaction between the exams and the semester was the outcome of interest. First, Levene's Test of equality of error variance showed that there was no violation of the assumption of homogeneity for my within-subjects factor (i.e., exams). The data did, unfortunately, violate the assumption of sphericity, according to Mauchly's Test (Mauchly's $W = .851$, $\chi^2(5) = 67.873$, $p < .001$, $\check{\varepsilon} = .900$; however, the closer the Greenhouse-Geisser calculated value ($\check{\varepsilon}$) is to 1, the closer the data are to being spherical; Field, 2009), so the F-ratio values that are reported below are those produced using the Greenhouse-Geisser correction to the degrees of freedom. That estimate value was chosen because its correction to the degrees of freedom used to evaluate the observed F-ratio is more conservative and appropriate in a case where there are four within-subjects conditions (Field, 2009). Because the data originally violated the assumption of sphericity, the multivariate test results should be reported as well, because the multivariate test does not require the data to be spherical (Field, 2009). The main effect of exam was significant (Pillai's Trace $V = .119$, $F(3, 421) = 18.978$, $p < .001$), as was the interaction between exam and semester, $V = .104$, $F(9, 1269) = 5.039$, $p < .001$. In terms of the linear analysis, there was a significant main effect of exams ($F(2.700, 1141.981) = 13.753$, $p < .001$, partial $\eta^2 = .031$), and a nonsignificant main effect for semester ($F(3, 423) = .594$, $p = .620$, partial $\eta^2 = .004$) . More importantly, there was a significant interaction between exams and semester, $F(8.099, 1141.981) = 4.331$, $p < .001$, partial η^2

= .030. Planned contrasts demonstrate that, in the context of the interaction between exams and semester, for the within-groups factor (i.e., exams), only the difference between the first and second exam was significant, $F(3, 423) = 10.860$, $p < .001$, partial $\eta^2 = .072$, (see Table 4.3), and the estimated means (see Table 4.4) reveal that the pattern of difference supports the hypothesis that the second exam scores would be higher than the first exam in the treatment semesters. However, the differences between the second and third exam, and the third and fourth exam are neither significant nor in a particular pattern that supports the hypothesis that scores on the later exams during treatment semesters would improve compared to the baseline semester (see Table 4.4, Figure 4.1). Rather, it seems that while Fall 2008 reflected the hypothesized pattern across all four exams, both spring semesters did not, though they did somewhat resemble each other. Post hoc pair-wise comparisons corroborated the null results of the semester main effect analysis.

Table 4.3
Repeated Exams Within-Subjects Contrast F-Test Results

Contrasts	F (3, 423)	Significance	Partial η^2
Exam 1 to Exam 2	10.860	.000	.072
Exam 2 to Exam 3	2.033	.109	.014
Exam 3 to Exam 4	.437	.727	.003



Note: The values represent the estimated mean after adjustment for the covariate.

Figure 4.1

Table 4.4

Means of Exams by Semester

Semester	Exams	Estimated Mean (*S.E.*)	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound	Descriptive Mean (*S.D.*)
Fall 2007 *N* = 96	1	38.164 (.728)	36.732	39.596	38.19 (7.320)
	2	38.160 (.734)	36.717	39.603	38.01 (7.006)
	3	37.815 (.848)	36.149	39.482	37.70 (7.135)
	4	38.852 (1.037)	36.813	40.891	38.66 (10.771)
	Total	38.25 (.690)	36.892	39.603	38.138 (6.417)
Spring 2008 *N* = 91	1	35.989 (.730)	34.554	37.424	36.09 (6.896)
	2	39.336 (.736)	37.889	40.782	39.37 (5.567)
	3	36.941 (.850)	35.271	38.611	36.99 (8.825)
	4	38.201 (1.040)	36.158	40.244	38.19 (10.350)
	Total	37.617 (.691)	36.259	38.975	37.659 (6.739)
Fall 2008 *N* = 124	1	35.445 (.641)	34.185	36.705	35.25 (5.967)
	2	39.716 (.646)	38.446	40.986	40.03 (7.250)
	3	39.544 (.746)	38.077	41.010	39.77 (6.381)
	4	39.644 (.913)	37.849	41.438	39.73 (8.096)
	Total	38.587 (.607)	37.394	39.780	38.695 (5.411)
Spring 2009 *N* = 132	1	38.666 (.612)	37.463	39.869	38.85 (7.049)
	2	39.524 (.617)	38.311	40.736	39.37 (7.178)
	3	37.974 (.712)	36.574	39.374	37.83 (8.794)
	4	38.787 (.872)	37.074	40.500	38.89 (9.552)
	Total	38.738 (.579)	37.599		

The nonsignificant results of the other comparisons offer neither support for the effect of the teams on exam performance, nor can retaining the null be called proof it does not exist, so the results reported above offer scant support for the hypothesis. While

scores for Exam 1 for the baseline semester were higher than two of the three treatment semesters, the second and third exams were clearly better than the baseline for the first two of the three treatment semesters. If the hypothesis had been fully supported, students in semesters including the teams would have shown increasing improvement throughout the semester compared to the baseline semester, hypothetically due to the increase in connection and utilization of the team for social and educational support. However, results of this analysis provided no concrete support for the prediction.

Finally, the prediction that students with more extra credit points for team studying contributions would have higher average exam scores than students with no or fewer team studying points was tested by regressing team studying bonus points from just students in semesters including teams on average exam scores (average across the four exams in the semester). The propensity score quintiles from the three semesters included in the analysis were also included in order to adjust the regression. Results supported the hypothesis. For every point of extra credit earned by contributing to their team's studying effort, students' average grade (not including bonus points) on the exams rose by three quarters of a point, $b = 0.792$, $SE = .132$, $p < .001$, adjusted $R^2 = .100$.

Effect of Teams on Students' Attitudes

The third set of hypotheses concerns students' attitudes toward the class, as measured on the course evaluation. First, I predicted that students in semesters including the teams would rate the class better than students would in the baseline semester. I tested this prediction by using an ANCOVA with semester as the IV and students' ratings in response to the question "Compared to other similar courses, how good was this class?" as the DV. Included in this analysis were the propensity score quintiles

(concerning students' year in school and major) as IVs, and the grade they expected to receive in the class as a covariate. The last variable - what grade they expected - was included because one might expect that students who thought that they did well in a class would consider it a better class than students who did poorly, just as a matter of course. I wanted to see specifically if I could attribute any difference more confidently to the actual impact of the innovation, independent of their perception of their own performance. Results partially supported the main hypothesis. In the covariate analysis of course rating, students' grade expectation significantly contributed to their opinion of the course, $F(1,285) = 8.589$, $p < .01$, partial $\eta^2 = .029$. As students' estimate of their grade moved higher (A was coded as 1), their opinion of the class improved, $b = -.302$, $p < .01$. There was a significant effect of semester on students' rating of the class's quality, $F(3,285) = 5.997$, $p < .01$, partial $\eta^2 = .059$. Planned contrasts revealed a nonsignificant difference between the baseline control semester and Spring 2008 and Fall 2008, and a strongly significant difference between the baseline and Spring 2009 (See Table 4.5). Fall 2008 and Spring 2009 were also significantly different from each other. Students in the last of the three treatment semesters rated the course as being better than students did in the control semester. Sidak-adjusted post hoc comparisons showed two significant differences among all of the semesters, with Fall 2007 and Fall 2008 being rated significantly less positively than Spring 2009. The reason that the post hoc analyses can report different results than the planned contrasts is because the post hoc analysis can be less powerful than the planned comparison (Field, 2009).

Table 4.5

Course Recommendation and Activities' Helpfulness by Semester

	Course Comparison		Activities' Helpfulness	
Semester	*M* (*SD*)	*M* (*SE*) Adjusted for Covariates	*M* (*SD*)	*M* (*SE*) Adjusted for Covariates
Fall 2007 (*N* = 68)	4.69 (1.44)	4.51 (.25)	4.97 (1.21)	5.15 (.23)
Spring 2008 (*N* = 64)	4.81 (1.37)	4.83 (.20)	5.00 (1.33)	4.85 (.18)
Fall 2008 (*N* = 84)	4.55 (1.52)	4.31 (.23)	5.06 (1.26)	5.02 (.20)
Spring 2009 (*N* = 102)	5.20 (1.37)	5.42 (.17)*	5.40 (1.15)	5.56 (.16)

** Significantly different from baseline semester at p<.05.*

I also predicted that students in semesters including the teams would rate the helpfulness of class activities higher than students would in the baseline semester. This prediction was tested using an ANCOVA with semester as the IV and students' rating of the helpfulness of the activities from the course evaluations as the DV. All four propensity score quintiles were included as IVs and expected grade was included as a covariate, as before. Results did not support the hypothesis. First, the covariate had a significant effect on students' ratings, $F(1,285) = 3.902$, $p < .05$, partial $\eta^2 = .014$, demonstrating that the higher that students rated the helpfulness of the activities, their estimated grade increased very slightly (A was coded as 1), though significantly, $b = -.183$, $p < .05$. The adjusted mean ratings students gave the class activities regarding their helpfulness toward understanding the material were not directionally supportive of the hypothesis; however, the overall effect of semester on students' ratings was significant, $F(3,285) = 3.483$, $p < .05$, partial $\eta^2 = .034$. The means (adjusted for the influence of the covariate) for each semester showed that during Spring 2008 and Fall 2008, students rated activities' helpfulness lower than during the baseline semester and Spring 2009, but

the differences were not significant according to the planned contrasts. Sidak-adjusted post hoc comparisons showed that the only significant difference among all of the semesters was between and Spring 2008 and Spring 2009.

I further tested this hypothesis using logistic regression with semester as the IV and students' yes or no response to the question, "Would you recommend this course to a friend?" as the DV to see any differences between semesters, while still controlling the same covariates. The combination of these two questions, Goodness and Recommendation, should be an apt indicator of the students' appreciation of the course in general. Results did not support the hypothesis, and the regression model was not good at predicting whether or not students would recommend the course to another student, –*2LL* = 334.393, *Cox and Snell* R^2 = 0.018, model $\chi^2(7, N = 318) = 5.809$, $p = .562$. See Table 4.6 for values and odds ratios for the predictors, all of which were nonsignificant.

Table 4.6
Logistic Regression Values for Course Recommendation by Semester

	95% CI for Odds Ratio		
	Lower	Odds Ratio	Upper
Fall 2007			
Spring 2008	.431	1.000	2.323
Fall 2008	.615	1.310	2.792
Spring 2009	.400	.871	1.896

Note: All values nonsignificant.

Effect of Peer Review on the Writing Assignment

The fourth set of hypotheses concerned the group writing assignment, and the difference between semesters when peer review was included in the grading scheme or not. The prediction that students would participate in the discussion portion of the

assignment more in semesters where peer review was used than when it was not was tested using an ANOVA with semester as the IV and students' scores on the discussion portion of the assignment as the DV. The reason that the sheer number of posts on the discussion board was not used in this analysis was because the posts' quality is a better indication of participation than just the number. For instance, a high number of poor quality posts resulted in a lower participation grade than a smaller number of high quality posts. Furthermore, any off-topic posts would artificially inflate the frequency information, but those were counted as off-topic in the grading scheme. The four propensity score quintiles were included as IVs in this analysis, which were calculated using information provided on the Student Information Sheet at the beginning of the semester, as described at length above. The reason I am reiterating this is to note that the propensity scores included whether students listed "group work" as a liked or disliked element of typical courses. In the analysis of this hypothesis, I relied on the balancing effect of the propensity score quintiles to correct any potential initial dissimilarity across semesters in terms of students' opinion of group work, even though there was not a significant difference across semester evident in the analysis that led to the creation of the propensity scores. However, theoretically, students' stated dislike for group work could have played a part in their participation in the project in later semesters, if they somehow knew to expect it as a course component. In other words, I had an a priori expectation that students' dislike of group work might confound comparison across semesters, which is the reason it was built into the propensity score calculation. The hypothesis was not supported. While results showed a significant effect of semester on students' participation grade for the writing assignment, $F(3,423) = 4.005$, $p < .01$, partial $\eta^2 =$

.028, the group means were not different in the hypothesized direction. Students in the baseline semester, during which no peer review was included, earned significantly higher grades in the participation component of the online writing assignment than students in all other semesters, including the last two semesters, which included peer review in the grading scheme (see Table 4.7). Planned contrasts and Sidak-adjusted post hoc pair-wise comparisons corroborated the finding that the only significant difference among all of the groups was between the grades during Fall 2007 and the three treatment semesters, meaning that there was no significant benefit of peer review revealed by this analysis.

Table 4.7

Discussion Participation Grade by Semester

Semester	Estimated Mean (*S.E.*)	95% Confidence Interval		Descriptive Mean (*S.D.*)
		Lower Bound	Upper Bound	
Fall 2007 *N* = 96	33.395 (1.128)	31.176	35.613	34.09 (11.823)
Spring 2008 *N* = 91	28.774 (1.131)	26.551	30.997	28.68 (10.795)
Fall 2008 *N* = 124	27.815 (.993)	26.626	30.530	28.73 (11.125)
Spring 2009 *N* = 132	28.578 (.948)	27.903	31.631	29.27 (9.117)

The prediction that the inclusion of the peer review element should be reflected in students' assessment of the writing assignment on course evaluations was tested using an ANCOVA with semester as the IV and writing assignment evaluation as the DV; the four propensity score quintiles were included as IVs, and students' expected grade in the class was included as a covariate. The hypothesis was not supported; there was no significant difference between students' appreciation of the writing assignment between the semesters that used peer review and semesters that did not, $F(3,306) = .812$, $p = .488$,

partial η^2 = .008. The covariate of students' expected grade was a marginally significant predictor of students' rating of the writing assignment, b = -.177, SE =.106 , p = .096, which can be interpreted as trending in the direction of students' rating of the writing assignment increasing the higher they anticipated their course grade to be (A was coded as 1). Students across the four semesters rated the writing assignment remarkably equally (see Table 4.8).

Table 4.8
Means of Students' Evaluation of Writing Assignment by Semester

Semester	Mean (*SD*)	*M* (*SE*) Adjusted for Covariates
Fall 2007 (*N* = 68)	5.25 (1.500)	5.37 (.27)
Spring 2008 (*N* = 64)	5.53 (1.480)	5.61 (.21)
Fall 2008 (*N* = 84)	5.49 (1.256)	5.49 (.24)
Spring 2009 (*N* = 102)	5.52 (1.621)	5.66 (.19)

Evaluation of the writing assignment rated on a 7-point Likert-type scale from A = Strongly Disagree, coded as 1 in the analysis, to G = Strongly Agree, with D = No opinion or Neutral as the midpoint, in response to the question "The written assignments allowed you to apply class ideas to your own ideas."
Note: All treatment semesters are not significantly different from the baseline.

Finally, the prediction that students in semesters including peer review would report a stronger sense of classroom community than students in semesters without peer review was tested using an ANCOVA with semester and the four propensity score quintiles as IVs, and the omnibus CCS score as the DV. The covariate was students' reported expected grade. This is the same analysis as was performed on the first hypothesis test reported at the beginning of the results section, but the focus on the results is shifted to compare the first two semesters to the last two semesters. Results did not support the hypothesis. Students' expected grade significantly contributed to their sense of overall classroom community, $F(1,285) = 11.138$, $p <. 01$, partial η^2 = .03 as a

covariate. As students' estimate of their grade moved higher (A was coded as 1), their sense of community increased, $b = -2.552$, $p < .01$. There was not a significant effect of semester on students' sense of community, $F(3,285) = 1.089$, $p = .354$, partial $\eta^2 = .011$, as reported above in the first section of the hypothesis tests. Directionally, results were slightly supportive, in that students in the three treatment semesters reported stronger sense of classroom community than the baseline control semester (see Table 4.1). However, Sidak-adjusted post hoc analysis revealed that there was no significant difference between the classroom community scores during Spring 2008 and the last two semesters. This pattern held across the omnibus CCS scale and both subscales, as reported above.

Chapter 5: Discussion

The results of the present study did not support the multiple hypothesized benefits of integrating this particular type of innovation into a large enrollment classroom. Contrary to expectations, students in semesters including teams did not experience a significantly stronger sense of community than students in the baseline semester. Both the overall sense of the classroom as well as the learning and connectedness subclassifications of the construct were not affected, though directional trends offer hope that the intervention was at least on the right track. The prediction that the home teams would improve students' performance on the online group writing assignment and unit exams on average was not supported. Only one of the semesters utilizing home teams showed significantly better later exams across the semester, compared to the baseline, though all treatment semesters' second exams were better than their and the baseline's first. Average exam scores were remarkably similar across all four semesters. The

students who contributed to their team studying efforts did indeed earn higher exam scores in general, as predicted. The prediction that students in semesters including teams would appreciate the course more, as indicated by their rating regarding the quality of the course and their willingness to recommend it to a friend, was partially supported in that students in the third treatment semester rated the course significantly higher than students in the baseline semester did, yet there was not a similarly significant difference between the other two treatment semesters and the baseline. Students' evaluation of the helpfulness of the class activities for their learning of the material was not linked as predicted to their sense of classroom community. The hypotheses that including peer review in the online writing assignment grade would increase students' participation in the assignment, their appreciation of the assignment, and their sense of classroom community were not supported.

While the few instances of nonsignificant but directional support among the many hypotheses are heartening, the results that did not confirm my expectations afford several areas where the home team construct would need to be modified in the future if it is to benefit students. Despite the present study's inconclusive results, the extant literature that led me to my hypotheses in the first place remains compelling. I proceed under the assumption that my largely null results are more indicative of my first three iterations of the intervention (the original and two replications) needing adjustment than the lack of a benefit to find.

The results of the present study corroborated some of the findings of previous research and conflicted with others. The team intervention did not produce a stronger sense of community in the classes utilizing them, except directionally, which is similar to

the marginally positive attitudes reported by students in McKinney and Graham-Buxton's (1993) and Lightner, Bober, and Willi's (2007) studies using cooperative learning groups. Delucchi (2006) found that learning outcomes measured objectively were improved by continued group work, but only later in the semester, and that result was directionally if not significantly present for one of my treatment semesters. While classroom community did not appear to be related to students' performance on exams in my study, McKinney, McKinney, Franuik, and Schweitzer (2006) and Rovai (2001) found that it was in their studies. On the other hand, students who worked cooperatively with their group to study for exams did earn higher exam scores. Rand, Dreber, Ellingsen, Fudenberg, and Nowak (2009) found that people who were rewarded for contributing unselfishly to group efforts contributed more to future group interactions than those not rewarded or punished for not contributing; one of my strongest significant effects was that students who contributed to their team's exam studying efforts (and earned bonus points for doing so) achieved higher exam scores than those who did not participate in group study activities. Machemer and Crawford (2007) found that students appreciated group activities which contributed to their exam success, and while I found students' contributions to be linked to improved exam scores, I did not find that their attitudes toward the class and its activities reflected the positivity Machemer and Crawford observed. In fact, Barfield's (2003) finding that students disliked group writing assignments was more similar to my results. Furthermore, contrary to what Cannon (2006) found, my students did not report the activities to be particularly helpful for their learning. Hijzen, Boekaerts, and Vedder (2006) reported that classroom community related positively to students' opinion of class activities, and I did not find that to be true in my study.

It is possible that the inability of the present study to fully support the hypotheses comes from multiple sources of error. First, it might be the case that the intervention, as a result of my attempt to retain a sense of normalcy and reduce disruption of the class, was not strong enough to elicit the widespread improvements I intended. Because a contrived or controlled situation differs from reality on several levels, experimental interventions need to be more extreme or intense than naturally occurring phenomena if differences are to be confidently identified. In my effort to preserve normalcy in an environment I was attempting to manipulate, I might have attenuated my ability to elicit or detect real differences. Perhaps the home teams need to be designed differently, or made more integral to the students' grade in the class in order to exert enough influence to change the overall learning and climate outcomes. If it is not the intervention itself that failed to produce the expected benefits, it might be that the measurement of its impact did not effectively identify the resulting differences. These ideas are explored in more depth below.

Strengths and Limitations of the Present Study

My first examination of this type of team innovation has elements that I consider strengths to recommend it and limitations to be addressed in future research. First, by replicating the treatment twice using slight variations and similar settings, this test of the innovation allowed me to explore the incremental evolution of the idea as I meshed theory with the reality of the classroom. This permitted me to make generalizations across the three semesters that would be weaker with only one semester to compare to the baseline. In my experience, any element of course curriculum design is a perpetual iterative process, and teachers who are systematic about the iterations, as I have been

here, are more likely to make consistent progress than those who go about the process with a random trial and error approach.

A limitation of this study is that this innovation has only been tested in the rather specific setting of a large-enrollment, intermediate-level social psychology class at a large state university in the Midwest, and therefore needs to be evaluated in multiple and different settings and on different populations of students before sweeping statements are supportable. The corollary strength of the constancy of the context of the three treatment semesters is that it controls for some potential situation-based variance. Ancillary to this point is that I taught this class in its large-enrollment format for the first time during the baseline semester. I had taught the same course material in smaller enrollment courses multiple times before, so the materials and policies were not novel to the experimental situation. However, because one might assume that there is a learning curve involved in switching to teaching a large-enrollment class, the results of the study could be confounded by my increasing comfort with the class size. Upon reflection, any difference in my behavior between the large and small format classes could probably be attributed to my ability to interact with individual students during class being hindered in a large class. Aside from that, my natural gain in confidence as an instructor over time is likely to have been different between my previous presentations of the course as small format classes and the large format classes. However, as all the classes involved in this study were the same style and size, the ways that my behavior differed between the smaller classes and the larger classes using the same materials is not necessarily relevant, but the history of the course as I have taught it should not be ignored when considering potential confounds in the present study.

This study employed a quasi-experimental design, which is at once both a strength and a limitation. It is a strength because it is the most realistic and ethical way to test a teaching method that is meant to be class-wide, semester-long, and beneficial. It is a limitation because causality can only be confidently determined in true experiments, so any conclusions about the effect of the innovation on the students' learning and sense of community have to be qualified. Further and extended tests of the idea would gradually allow for convergent evidence to support conclusions drawn with more confidence. That does not entirely make up for the lack of the ability to infer causality, but does improve external validity.

The particular threats to internal validity that quasi-experimental research must consider are participants' experiences between measurements (i.e., history), participant maturation (i.e., changes occurring naturally due to time passing), re-testing effects, changes in instrumentation, regression to the mean, selection biases, mortality (i.e., attrition), and an interaction between participant selection and maturation (Campbell & Stanley, 1963). The NEGD is not typically vulnerable to history, maturation, re-testing, instrumentation, selection, or mortality threats. My use of the same course material across semesters except for the intervention, as mentioned above, avoids instrumentation effects. The design can be vulnerable to regression to the mean effects when relevant, and researchers using NEGD should attend to the risk of a selection-maturation interaction (i.e., people maturing differently in different groups; Campbell & Stanley, 1963). The regression to the mean threat to internal validity for quasi-experimental designs results from non-random sampling of participants from the extremity of distributions on some relevant evaluation, typically due to interest in them because of

their extreme scores (e.g., choosing subjects with the highest scores on a measurement). This threat is not present for this study because my students were not chosen to be in a particular class based on any standardized criteria (*all* were required to have taken the Introductory Psychology prerequisite or equivalent). My study's use of three treatment groups to compare to the control makes it unlikely that any one intervention semester's students' potential history, maturation, or selection effects could have been mistakenly identified as a treatment effect (in effect, using replication to test reliability). The first time the intervention was used in the class provided a semester in which none of the students could have possibly known about the intervention and have chosen to enroll for that specific aspect. If the control group was in some way initially different from the three treatment groups, then I would have expected the final differences observed between it and each of the three treatment groups were similar (meaning the treatment groups resembled each other, as opposed to the observed outcome difference between one of the treatment groups and the control group being different from the observed outcome differences between the control group and the other two treatment groups).

Finally, I designed the implementation of the innovation and its subsequent evaluation to be as non-intrusive as possible during the classes included in this investigation, in order to maintain the sense of normality expected by the students in the class. If they felt like they were guinea pigs for a semester, as if they had to act a certain way to uphold (or derail) my expectations (an internal validity threat generalized as *demand characteristics*), or that their inclusion in this research somehow affected their grade in the class, the results would be tainted both methodologically and ethically. Any perception of coercion, artificiality, or subterfuge on the students' part limits the internal

validity of the study. Therefore, the choices I made about how to measure the constructs of interest were in the context of controlling for the possible internal validity threats inherent to quasi-experimental research, particularly threats from practice effects, re-testing effects, changes in instrumentation, demand characteristics, and in this particular study, the mere fact that observation itself can change people's behavior (a phenomenon called the *Hawthorne effect*; Parsons, 1974).

Recommendations for Future Research

The process of determining best practices for any real life, and therefore complex, classroom situation or learning objective is fundamentally iterative, and likely to be interminable (in a good way). Future investigations of this type of intervention or innovation would need to pursue a couple of different avenues. One direction would be to find new or different ways to measure classroom community. In the present study, the measurement of students' sense of community was attached to the course evaluations in order to capitalize on that standard and expected event without alerting students to its separate aim. Consequently, a student's CCS scores could not be matched with his or her other class records, such as his or her grade in the class, limiting the evaluation to cross-semester aggregates only. A specific strategy to ameliorate this problem would be to link an extra-curricular content exam to CCS scores. By appropriating a procedure used in many departments for program assessment, a researcher could link specific content knowledge acquisition to the measurement of classroom community per student. In this approach, a researcher would administer a brief content exam on the first day of class, with the announcement that students' scores on the test would be completely separate from their course grade. Students' need not worry about doing poorly on the

exam (which is a cumulative final, for all intents and purposes) because they have not yet experienced the course and are expected not to do well. Presented as a "test of the teacher" rather than a test of their knowledge, students are likely to be less anxious or disgruntled about a surprise exam on the first day of class. Students retake the exact same exam again on the last day of class, and both pre-test and post-test are linked to students' scores on the sense of classroom community measurement. Neither the content exam nor the CCS scores are officially affiliated with students' scores in the class (i.e., the content exam does not supplant a regular course exam), and so is not as confounded with other course grade elements or considerations. Course evaluations would still be separate from this process in order to protect students' anonymity.

Furthermore, a direct measurement of change in knowledge and sense of community would be linked for each individual and control for students' varying levels of pre-knowledge of psychology. This association between the students' content learning and their sense of classroom community is necessary to investigate the potential individual level relationships between subjective climate and objective learning outcomes. The pretest-posttest approach will also allow delineations between high achieving students' and their possible pre-existing inclination to feel more positive about the classroom community. The present study did not have the ability to investigate fine-grain measurements in students' performance in the writing assignment, for example, in pursuit of specific instructional objectives. With the subjective experience linked more explicitly with their performance measurements, I could see where the intervention either is particularly effective, or identify more specific obstacles to meeting my instructional objectives and learning goals.

Second, there is not a good pre-test for classroom community that avoids the potential change in behavior students might exhibit if they become aware of being observed for classroom community. On the first day of a class, it is patently ridiculous to ask students how much they feel this particular classroom climate is conducive to their learning, because they have yet to experience whether their interactions with classmates and the teacher are positive or negative. People, not just students, are not typically adept at predicting their future emotional states while accounting for all of the other extenuating and mitigating factors that directly influence those emotions (see literature on affective forecasting, Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998), so asking them to predict their sense of community in a class on the first day holds little merit. Therefore, other non-suspicious proxies for classroom community must be used if a pre-test of that construct is desired. A possible alternative would be to ask a separate class's students to report their sense of community for similar classes (by type or size) they had taken in the past and use it to anecdotally compare to experimental participants' report at the end of the semester.

On the other hand, students' sense of social identity within their teams is not necessarily being measured with the semester-end measurement of classroom community. That instead might be more indicative of their sense of the entire class as a community. If students' collective identity within their teams was directly and explicitly measured over the course of the semester (e.g., at midterm and finals), their performance measures in exams and assignment grades could be more clearly investigated in terms of students' investment and engagement with their teams. This approach would require measurement of identity to be linked to their course records, and would also afford an

opportunity to investigate the composition of the teams (i.e., their members' similarity or difference in demographic variables) to identify the effect of teams on various *types* of students.

I would characterize the scholarly approach to teaching and learning as a two-step sequence, with description and understanding of a situation necessarily preceding manipulation of it. Therefore, another avenue to pursue in future research is to examine the social and cognitive processes underlying a connection between classroom community, learning, and inclusion of a team system in formal course components. The present study was looking for the *existence* of the hypothesized effect, and the innovation design was based on an aggregation of group composition recommendations gleaned from social psychology, in essence beginning the trek toward best practices by starting with best guesses.

The home team innovation examined here in its first permutation incorporated elements designed to promote group identity (i.e., the team names, mottos, mascots, and photographs), common goals (i.e., earning extra credit by performing the best as a team on exams), frequent interaction for a meaningful length of time (i.e., multi-modal communication throughout the semester), out-group competition with other teams in the class, and identifiable individual contributions to the team's overall outcome (i.e., individual rewards for personal contributions to team studying efforts). Further investigation of these elements could explore the progression of identity formation within the team, by measuring their sense of group identity over the course of the semester. Or, a study could examine the effect of making the shared goals for collaborative team success more overt and explicit in terms of the students' contributions to the team

studying effort. Even though having students formally plan study group team events, write meeting objectives or agendas, and report on group progress toward the stated goals would increase their out-of-class work load and therefore potentially create resistance, it would be a more direct measurement of the interdependence component of group identity. (A possible downside of this strategy could be that it takes away from individual study time and could therefore actually hinder performance.) If a teacher created review days in class using some of the game paradigms that are popular (e.g., Jeopardy), the competition between teams in the class could be made more salient as well. Both of these ideas would also increase frequent meaningful interaction among team members. Future research should weigh these suggestions against the considerations of risk of students' perception of coercion and work load for both teachers and students, but there are benefits to finding more information on the underlying process of group function and identity. There are myriad different directions from which to approach the relationship between pedagogy and social psychology, and therefore a wealth of opportunity to improve teaching in the college classroom.

In due course, further testing of the overarching hypothesis, that incorporating teams in the classroom improves students' learning and sense of classroom community, will be useful as a next step in improving the large-enrollment classroom's questionable effectiveness. This first examination of my home team innovation attempted to manipulate some of the group dynamics that can influence performance. As many different ways as there are to incorporate group work into a course, the strategy tested in the present study is certainly just one way that the theoretical integration of social psychology with the scholarship of teaching should be explored.

Implications and Conclusions

Three uses of the results of this study, despite the inconclusive results, are particularly desirable. First, teachers can consider the particular innovation investigated in this study and apply their own insight or background to improve it for use in their own classrooms to help them meet their own classes' learning goals and instructional objectives. This strategy should be applicable in any classroom regardless of subject matter. Furthermore, if the intervention is scaled to match the size of the class in which it is being implemented, it could still hold value whether it is incorporated into a class with an enrollment of 30 or 300. While I intended the intervention to emulate a small class environment and climate, nothing precludes this approach being used in small classes as well as large.

Second, teachers can improve their other lesson plans by using the *rationale* demonstrated herein, by applying social psychological principles to their teaching methods. This will be most relevant to social psychology faculty of course, but other psychologists with Introductory Psychology-level understanding of social psychology principles could find valuable pieces of research and theory to integrate into their lesson planning. If non-psychology faculty are to benefit from this, teachers who do successfully use social psychology to improve course components in terms of students' learning and classroom management need to widely distribute their findings. Publication in Scholarship of Teaching and Learning (SoTL) journals is better for reaching an interdisciplinary audience than specifically publishing in teaching of psychology journals, and presentations at conferences that are non-subfield-specific, interdisciplinary, or national are more useful than keeping results in teaching-specific divisions of scholarly

societies. On a local level, if psychologists using this approach will participate in their campuses' professional development efforts or teaching and learning centers, they will reach other faculty on campus, and even invite opportunities for interdisciplinary collaboration. This initiative to integrate the two disciplines would particularly benefit from administrations placing similar value on research on teaching as there is research on discipline content in terms of faculty evaluation for promotion and tenure.

Finally, students can see the usage of applied psychological research and theory in their classrooms and have both the personal experience to aid their understanding of the principles, as well as the knowledge that their teachers are practicing what they teach. If illustrations of applied psychology are the objective, the teachers should be explicit and transparent about the purpose, motivation, and empirical support behind integrations of targeted interventions, or else run the risk of students feeling as if they are jumping through arbitrary hoops.

Integrating SoTL and social psychological theory can also benefit social psychology as a field. Social psychology is very broadly applicable (and testable) across social situations, but tends to be investigated experimentally in more limited venues. Social psychologists can work toward strengthening their explanations of social phenomena by including a new, unique context in which to examine them. Laboratory studies and field observations offer the perennial tradeoff between control and realism. Testing social psychological theory in the classroom, an environment both controllable and natural, can strengthen the theoretical underpinnings of social psychological doctrine. Replications of studies that have formed the foundation of the subfield are not always attractive projects to pursue, especially in light of the demand for novelty in the

publication process, but replication with extension to a new environment or a new operationalization of a construct is still critical to our ability to progress as a field of study. Because this theoretical integration with SoTL also has value as a practical application of social psychology, it should be part of social psychology's scientific process.

While there are multiple proposed explanations in the literature for why large-enrollment classes might not be ideal for the kind of critical thinking and long-term content and skill acquisition universities are working to promote, some of the shortcomings of large, lecture-based classes can be addressed by designing active learning alternatives using social psychological group theory. The particular problems being faced in large-enrollment classrooms that might be mitigated by applying social psychological theory include: the growing sense of anonymity or isolation many students in large classes report feeling (Messino, Gaither, Bott, & Ritchey, 2007), the burden of responsibility for students' learning being placed solely on the teacher's shoulders (Lightner, Bober, & Willi, 2007; Machemer & Crawford, 2007; Smith, 1996), the slackening of enthusiasm for learning and student attrition documented in university populations (Dawson, 2006; Gupta, 2004; Healey & Matthews, 1996; Machemer & Crawford, 2007), and students' own diminishing expectation of success (Johnson & Johnson, 1994; Long & Coldren, 2006), along with other practical and emotional influences on student learning.

My reactions to the present study's outcome does somewhat resemble the conclusions Delucchi (2006) drew in his examination of collaborative learning strategies in his statistics course. He concluded his report by saying that despite his intervention

not succeeding in all the ways he had hoped as first tested, he would continue to incorporate an iteratively revised version of it in future courses, because the problem he had attempted to address with a group work intervention persisted. As disappointing as some of the results of this first examination of my intervention are, I too am still hopeful that there is a way to increase the improvements of the learning outcomes and classroom climate in large classes, and I remain convinced that social psychology offers useful insight for that process. As a microcosm of society, the college classroom is a unique environment; this may seem to limit the generalizability of the information gleaned from this study and future research on this particular approach to teaching. However, for students who represent the future of this society, the importance of high quality teaching and course design to cannot be understated. Therefore, the application of social psychological theory and principles to the psychology of teaching and learning is a practical and valuable pursuit.

References

Aronson, E., & Bridgeman, D. (1979). Jigsaw groups and the desegregated classroom: In pursuit of common goals. *Personality and Social Psychology Bulletin, 5,* 438-466.

Austin, P. C. (2011). A tutorial and case study in propensity score analysis: an application to estimating the effect of in-hospital smoking cessation counseling on mortality. *Multivariate Behavioral Research, 46,* 119-151.

Barfield, R. L. (2003). Students' perceptions of and satisfaction with group grades and the group experience in the college classroom. *Assessment & Evaluation in Higher Education, 28,* 355-369.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (ed.), *Handbook of research on teaching*. Chicago: Rand McNally.

Cannon, P. (2006). Enhancing understanding and interest through group discussion. *College Teaching, 54,* 211.

Carnes Stevens, M. (2007). Making groups work. *College Teaching, 55,* 88.

Chapman, J. G., Arenson, S., Carrigan, M. H., & Gryckiewicz, J. (1993). Motivational loss in small group tasks: Free-riding on a cognitive task. *Genetic, Social, and General Psychology Monographs, 119,* 57-73.

Chapman, L., & Ludlow, L. (2010). Can downsizing college class sizes augment student outcomes? An investigation of the effects of class sizes on student learning. *The Journal of General Education, 59,* 105-123.Cohn, C. L. (1999). Cooperative learning in a macroeconomics course. *College Teaching, 47,* 51-54.

Dawson, S. (2006). A study of the relationship between student communication interaction and sense of community. *Internet and Higher Education, 9,* 153-162.

Delucchi, M. (2006). The efficacy of collaborative learning groups in an undergraduate statistics course. *College Teaching, 54,* 244-248.

Evans, C. R., & Dion, K. L. (1991). Group cohesion and performance: A meta-analysis. *Small Group Research, 22,* 175-187.

Exeter, D. J., Ameratunga, S., Ratima, M., Morton, S., Dickson, M., Hsu, D., & Jackson, R. (2010). Student engagement in very large classes: The teachers' perspective. *Studies in Higher Education, 35,* 761-775.

Fassinger, P. A. (2000). How classes influence students' participation in college classrooms. *Journal of Classroom Interaction, 35,* 38-47.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology, 58,* 203-211.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed). London: Sage Publications.

Fox-Cardamone, L., & Rue, S. (2003). Students' responses to active-learning strategies: An examination of small-group and whole-class discussion. *Research for Educational Reform, 8,* 3-15.

Gaertner, S. L., Dovidio, J. F., Banker, B. S., Houlette, M., Johnson, K. M., & McGlynn, E. A. (2000). Reducing intergroup conflict: From superordinate goals to decategorization, recategorization, and mutual differentiation. *Group Dynamics: Theory, Research, and Practice, 4,* 98-114.

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 75,* 617-638.

Gray, T., & Madson, L. (2007). Ten easy ways to engage your students. *College Teaching, 55,* 83-87.

Gupta, M. L. (2004). Enhancing student performance through cooperative learning in physical sciences. *Assessment & Evaluation in Higher Education, 29,* 63-73.

Hahs-Vaughn, D. L. & Onwuegbuzie, A. J. (2006). Estimating and using propensity score analysis with complex samples. *The Journal of Experimental Education, 75,* 31-65.

Harkins, S. G., & Jackson, J. M. (1985). The role of evaluation in eliminating social loafing. *Personality and Social Psychology Bulletin, 11,* 457-465.

Harkins, S. G. (1987). Social loafing and social facilitation. *Journal of Experimental Social Psychology, 23,* 1-18.

Harkins, S. G. (2006). Mere effort as the mediator of the evaluation-performance relationship. *Journal of Personality and Social Psychology, 91,* 436-455.

Hart, G. (1975). Preparing students for small group work in the classroom. *Education, 95,* 351-353.

Healey, M., & Matthews, H. (1996). Learning in small groups in university geography courses: Designing a core module around group projects. *Journal of Geography in Higher Education, 20,* 167-181.

Hijzen, D., Boekaerts, M., & Vedder, P. (2006). The relationship between the quality of cooperative learning, students' goal preferences, and perceptions of contextual factors in the classroom. *Scandinavian Journal of Psychology, 47,* 9–21.

Imbens, G. W. (2000). The role of propensity score in estimating dose-response functions. *Biometrika, 87,* 706-710.

Jackson, J. M., & Williams, K. D. (1985). Social loafing on difficult tasks: Working collectively can improve performance. *Journal of Personality and Social Psychology, 49,* 937-942.

Johnson, D. W., Johnson, R. & Smith, K. (1991). *Cooperative learning: Increasing college faculty instructional productivity*. Washington, DC: The George Washington University, Graduate School of Education and Human Development.

Johnson, D. W., Johnson, R., & Stanne, M. B. (2000). Cooperative learning methods: A meta-analysis. *Journal of Research in Education, 12*, 5-14. Retrieved from http://www.co-operation.org/pages/cl-methods.html.

Johnson, R., & Johnson, D. W. (1994). An overview of cooperative learning. In J. Thousand, R. Villa, & A. Nevin (Eds.), *Creativity and collaborative learning* (pp. 31-44). Baltimore: Paul Brooks Publishing. Retrieved from http://www.co-operation.org/pages/overviewpaper.html.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65*, 681-706.

Kerr, N. L. (1983). Motivation losses in small groups: A social dilemma analysis. *Journal of Personality and Social Psychology, 45*, 819-828.

King, P. E., & Behnke, R. R. (2005). Problems associated with evaluating student performance in groups. *College Teaching, 53*, 57-61.

Kreie, J., Headrick, R. W., & Steiner, R. (2007). Using team learning to improve student retention. *College Teaching, 55*, 51-56.

Lancy, D. F., & Rhees, A. (1994). A sense of community. *College Teaching, 42*, 102-107.

Langfred, C. W. (1998). Is group cohesiveness a double-edged sword? *Small Group Research, 29*, 124-144.

Latane, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37*, 822-832.

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology, 28*, 129–137.

Levine, J. M., & Moreland, R. L. (1998). Small groups. In Gilbert, D. T., Fiske, S. T., & Lindzey, G. (Eds.), *The handbook of social psychology*, (4th ed., pp. 415-469). New York: McGraw Hill.

Lightner, S., Bober, M. J., & Willi, C. (2007). Team-based activities to promote engaged learning. *College Teaching, 55*, 5-18.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation. *American Psychologist, 57*, 705-716.

Locke, E. A., & Latham, G. P. (2006). New directions in goal-setting theory. *Current Directions in Psychological Science,15*, 265-268.

Long, H. E., & Coldren, J. T. (2006). Interpersonal influences in large lecture-based classes: A socioinstructional perspective. *College Teaching, 54*, 237-243.

Machemer, P. L., & Crawford, P. (2007). Student perception of active learning in a large cross-disciplinary classroom. *Active Learning in Higher Education, 8*, 9-30.

McKinney, J. P., McKinney, R. F., Franiuk, R., & Schweitzer, J. (2006). The college classroom as a community: Impact on student attitudes and learning. *College Teaching, 54*, 281-284.

McKinney, K., & Graham-Buxton, M. (1993). The use of collaborative learning groups in the large class: Is it possible? *Teaching Sociology, 21*, 403-408.

Messino, M., Gaither, G., Bott, J. & Ritchey, K. (2007). Inexperience versus experienced students' expectations for active learning in large classes. *College Teaching, 55*, 125-133.

Monk-Turner, E., & Payne, B. (2005). Addressing issues in group work in the classroom. *Journal of Criminal Justice Education, 16*, 166-179.

Mullen, B., & Copper, C. (1994). The relation between group cohesiveness and performance: An integration. *Psychological Bulletin, 115*, 210-228.

Mulryan-Kyne, C. (2010). Teaching large classes at a college and university level: Challenges and opportunities. *Teaching in Higher Education, 15*, 175-185.

Parsons, H. M. (1974). What happened at Hawthorne? Science, 183, 922-932.

Pedder, D. (2006). Are small classes better? Understanding relationships between class size, classroom processes, and pupils' learning. *Oxford Review of Education, 32*, 213-234.

Pickett, C. L. (2001). The effects of entitativity beliefs on implicit comparisons between group members. *Personality and Social Psychology Bulletin, 27*, 515-526.

Pimmel, R. J. (2003). A practical approach for converting group assignments into team projects. *IEEE Transactions on Education, 46*, 273-282.

Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D., & Nowak, M. A. (2009). Positive interactions promote public cooperation. *Science, 325*, 1272-1275.

Rovai, A. P., & Wighting, M. J. (2005). Feelings of alienation and community among higher education students in a virtual classroom. *Internet and Higher Education, 8*, 97-110.

Rovai, A. P. (2001). Building classroom community at a distance: A case study. *Educational Technology, Research, and Development, 49*, 33-48.

Rovai, A. P. (2002a). A preliminary look at the structural differences of higher education classroom communities in traditional and asynchronous learning network courses. *Journal of Asynchronous Learning Networks, 6*, 41-56.

Rovai, A. P. (2002b). Development of an instrument to measure classroom community. *Internet and Higher Education, 5*, 197-211.

Rovai, A. P. (2007). Facilitating online discussion effectively. *Internet and Higher Education, 10*, 77-88.

Shepperd, J. A. (1993). Productivity loss in performance groups: A motivation analysis. *Psychological Bulletin, 113*, 67-81.

Sherif, M. (1956). Experiments in group conflict. *Scientific American, 195*, 54-58.

Silvia, P. J. (2005). Deflecting reactance: The role of similarity in increasing compliance and reducing resistance. *Basic and Applied Social Psychology, 27,* 277-284.

Smith, K. A. (1996). Cooperative learning: Making 'group work' work. In Sutherland, T. E., and Bonwell, C. C. (Eds.), *Using active learning in college classes: A range of options for faculty*, San Francisco: Jossey-Bass.

Steinbrink, J. E., & Jones, R. M. (1993). Cooperative test-review teams improve student achievement. Clearing House, 66, 307-312. Retrieved from Academic Search Premier.

Summers, J. J., & Svinicki, M. D. (2007). Investigating classroom community in higher education. *Learning and Individual Differences, 17,* 55-67.

Summers, J. J., Beretvas, S. N., Svinicki, M. D., & Gorin, J. S. (2005). Evaluating collaborative learning and community. *The Journal of Experimental Education, 73,* 165-188.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behavior. *European Journal of Social Psychology, 1,* 149-178.

Urdan, T., & Schoenfelder, E. (2006). Classroom effects on student motivation: Goal structures, social relationships, and competence beliefs. *Journal of School Psychology, 44,* 331-349.

Vesilind, P. A. (2000). *So you want to be a professor? A handbook for graduate students.* Thousand Oaks, CA: Sage Publications.

Vik, G. N. (2001). Doing more to teach teamwork than telling students to sink or swim. *Business Communication Quarterly, 64,* 112-19.

Williams, K. D., & Karau, S. J. (1991). Social loafing and social compensation: The effects of expectations of co-worker performance. *Journal of Personality and Social Psychology, 61,* 570-581.

Wolfe, K., Lee, K. E., Wu, C. M. E., & Gould, R. (2003). An assessment of self-selected versus instructor-assigned teams in an introductory tourism course. *Journal of Teaching in Travel & Tourism, 3(3),* 1-18.

Woo, Y., & Reeves, T. C. (2007). Meaningful interaction in web-based learning: A social constructionist interpretation. *The Internet in Higher Education, 10,* 15-25.

Yanovitzky, I., Hornik, R., & Zanutto, E. (2008). Estimating causal effects in observational studies: The propensity score approach. In A. F. Hayes, M. D. Slater, L. B. Snyder, A. F. Hayes, M. D. Slater, L. B. Snyder (Eds.), *The Sage Sourcebook Of Advanced Data Analysis Methods For Communication Research* (pp. 159-184). Thousand Oaks, CA US: Sage Publications, Inc.

Zajonc, R. B. (1965). Social facilitation. *Science, 149,* 269-274.

Zanutto, E., Lu, B., & Hornik, R. (2005). Using propensity score subclassification for multiple treatment doses to evaluate a national antidrug media campaign. *Journal of Educational and Behavioral Statistics, 30,* 59-73.

Appendix A

Student Information Sheet

PSYC 288 Psychology of Social Behavior
Spring 2009
Instructor: Bethany Johnson

Student Info Sheet

Name: _____

Circle your year in school: First-year Sophomore Junior Senior Other

Are you planning to graduate in May '09? ___No ___Yes

Major: _____

Email address: _____

Phone: _____

****
******

Do you have a job outside of school? ___No ___Yes (average hrs/week _____)

Number of hours enrolled in during this semester: _____hrs

Psychology classes already taken:

Reason for taking Social Psychology:

Things you like in classes (for example, lectures, group activities, discussion, movies, etc.):

Things you don't like in classes (for example, lectures, group activities, discussion, movies, etc.):

Something interesting about me is:

Appendix B

List of Movies for Online Group Writing Assignment

The following are the movies from which groups choose for the online discussion assignment. Each group is randomly given three movies, and they decide together which one they want to watch. Because there are generally about 22 groups in a class, each of the movies appears about three times.

1. A History of Violence (2005) Viggo Mortenson
2. American Beauty (1999) Kevin Spacey
3. Being John Malkovich (1999) John Cusack
4. Being There (1979) Peter Sellers
5. Bend It Like Beckham (2002) Keira Knightly
6. Boys Don't Cry (1999) Hillary Swank
7. Brokeback Mountain (2005) Heath Ledger
8. Chocolat (2000) Juliette Binoche
9. Crash (2005) Paul Haggis, director
10. Das Leben der Anderen ("The Lives of Other People") (2006) Martina Gedeck
11. Dead Poets' Society (1989) Robin Williams
12. Eternal Sunshine of the Spotless Mind (2004) Jim Carrey
13. Footloose (1984) Kevin Bacon
14. Good Will Hunting (1998) Matt Damon
15. Juno (2007) Ellen Page
16. Monty Python and the Holy Grail (1975) John Cleese
17. Rashomon (1950) Akira Kurosawa, director
18. Shall We Dance (1996) Masayuki Suo, director
19. The Princess Bride (1987) Robin Wright-Penn
20. The Usual Suspects (1994) Kevin Spacey
21. The Shawshank Redemption (1994) Tim Robbins
22. Whale Rider (2002) Kiesha Castle-Hughes

Appendix C

Course Evaluations including Class Community Scale Addendum

Course Evaluation Form
Introduction to Social Psychology 288

Instructor: Bethany Johnson

Spring 2009

Note: Do not put your name anywhere on this evaluation.
The instructor will not see these evaluations until after grades have been submitted.

Bubble in the letter corresponding to your rating for each statement that appears below.

A	**B**	**C**	**D**	**E**	**F**	**G**
strongly disagree	disagree	somewhat disagree	neutral or no opinion	somewhat agree	agree	strongly agree

1. Class meetings began and ended on time and at the scheduled/arranged time.
2. The lectures or other class presentations were clear and well-organized.
3. It was easy to take notes on the lectures.
4. The instructor knew if the class was understanding her or not.
5. The instructor had an interesting presentation style.
6. The presentation style of the instructor was consistent.
7. The instructor appeared to be knowledgeable and up-to-date in the subject.
8. The instructor's answers to questions were understandable and to the point.
9. The instructor was respectful of diverse points of view and opinions.
10. The instructor maintained an atmosphere that supported the expression of ideas by students.
11. The exams appropriately covered readings and classroom presentations.
12. The instructor was available to students during scheduled office hours.
13. The instructor treated students fairly and without regard to personal characteristics (e.g., gender, ethnicity, political views, orientation, etc.)
14. The course led you to engage in active thinking about the subject or its application to real-world issues.
15. The in-class exercises encouraged you to think about and apply the class material to real-world issues.

16. The written assignments allowed you to apply class ideas to your own ideas.
17. The course's BlackBoard site was useful and easy to navigate.
18. The course's textbook was useful and easy to read.

For the next three questions, use the following scale:

A	**B**	**C**	**D**	**E**	**F**	**G**
among the worst	a lot worse than average	a little worse than average	average	a little better than average	a lot better than average	among the best

19. Compared to other instructors you've had, how good was this instructor?
20. Compared to other courses you've taken at this level, how good was this course?
21. Compared to other textbooks you've used, how good was this textbook?

22. Your year in school: A=First-year B=Sophomore C=Junior D=Senior E=Other
23. Are you a psychology major? A = yes B = no
24. What grade do you expect to receive in this class? (Bubble in the letter grade on your answer sheet)
25. Would you recommend this course to another student? A = yes B = no
26. Would you recommend this person as an instructor to another student? A = yes B = no

---

For the questions below, please write out your answers neatly.

What did you like <u>most</u> about the class?

What did you like <u>least</u> about the class?

What would you like the instructor to know about her teaching content, style, or approach?

What did you like about the exams? What would you change?

What did you like about the writing assignments? What would you change?

Do you have any other comments about the course or the instructor?

For the next four questions, use the following scale:

A	**B**	**C**	**D**	**E**	**F**	**G**
Completely confusing	Pretty confusing	A little bit confusing	Neutral or no opinion	A little bit helpful	Nice and helpful	Extremely helpful

27. Rate the helpfulness of the LECTURES to understanding the material.
28. Rate the helpfulness of the ACTIVITES to understanding the material.
29. Rate the helpfulness of the CHAPTERS to understanding the material
30. Rate the helpfulness of the ARTICLES to understanding the material.

For the next twenty questions, use the following scale:

A	**B**	**C**	**D**	**E**
strongly disagree	disagree	neutral or no opinion	agree	strongly agree

31. I feel that students in this course care about each other.
32. I feel I am encouraged to ask questions.
33. I feel connected to others in this course.
34. I feel that it is hard to get help when I have a question.
35. I do not feel a spirit of community in this class.
36. I feel that I receive timely feedback.
37. I feel that this class is like a family.
38. I feel uneasy exposing gaps in my understanding.
39. I feel isolated in this course.
40. I feel reluctant to speak openly.
41. I trust others in this course.

42. I feel that this course results in only modest learning.
43. I feel that I can rely on others in this course.
44. I feel that other students do not help me learn.
45. I feel that members of this class depend on me.
46. I feel that I am given ample opportunities to learn.
47. I feel uncertain about others in this class.
48. I feel that my educational needs are not being met.
49. I feel confident that other students will support me.
50. I feel that this course does not promote a desire to learn.