8-2015

The Effects of Voluntary versus Cold-calling Participation on Class Discussion and Exam Performance in Multiple Sections of an Educational Psychology Undergraduate Course

Brittany Ann Carstens
University of Tennessee - Knoxville, bcarste1@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Brittany Ann Carstens entitled "The Effects of Voluntary versus Cold-calling Participation on Class Discussion and Exam Performance in Multiple Sections of an Educational Psychology Undergraduate Course." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in School Psychology.

Robert L. Williams, Major Professor

We have read this dissertation and recommend its acceptance:

Sherry K. Bain, Dennis J. Ciancio, David F. Cihak

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

The Effects of Voluntary versus Cold-calling Participation on Class Discussion and Exam Performance in Multiple Sections of an Educational Psychology Undergraduate Course

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Brittany Ann Carstens
August 2015

Acknowledgements

There are several people I would like to take this opportunity to thank. First, I would like to thank my committee members for all their insight and help on not only my dissertation, but also throughout my graduate career. Dr. Robert Williams, my advisor and committee chair, has been an invaluable mentor. His strong work ethic, attention to detail, and dedication to pushing his students further have helped to develop my skills as both a researcher and a practitioner. Dr. Sherry Bain has continually taken the time to aid in the development of my professional skills. Her help in advancing my writing skills has made me a more effective writer. Dr. Dennis Ciancio was a great help with data analyses and was always willing to take the time to discuss methodological design and possible methods of approach. Dr. David Cihak has provided valuable feedback on methodological design and disseminating the findings of research so it may be useful in a real world application. I am very grateful for each of these members and the unique contribution each has made to my professional growth and development. I would also like to thank the School Psychology faculty for all the knowledge and skills they have imparted to me across the last several years.

I am indebted to my fellow research team members. Thank you to the instructors of Educational Psychology 210 at the time of this study: Tiffany Best, Katie Crabtree, Leslie Hart, and Ellie Trant. It was your dedication and commitment to learning and implementing the ins and outs of this study that made it possible. I greatly appreciate your insight into the real-world application of cold-calling and all the hard work you invested to make my study successful. I would also like to thank both Samantha Adair and Carrie Jaquett. These two remarkable graduate assistants aided in data collection and offered helpful feedback throughout the study. I owe a special thanks to Drs. Carolyn Blondin and Charles Galyon. They are both always willing

to bounce ideas around with me and provide support when needed. I am so thankful to all of these individuals for their amazing support and hard work.

I would like to thank my family, who has always been encouraging and supportive of all my endeavors. Thank you to my mother for unfailingly making herself available when needed whether for moral support or professional advice. Thank you to my father for helping me to see the lighter side of both graduate school and life. Thank you to Dr. Briana Fiser, my sister, for traversing graduate school first and providing me someone to share all the ups and downs that come with it. Lastly, I would like to thank my husband Jason Wilder for helping me to find my balance. I am so grateful to have the love and support of these wonderful people.

Abstract

Although class participation has been linked to improved student performance, little research has evaluated the effects of cold-calling versus voluntary participation. This study (*N* =156) determined the differential effects of voluntary and cold-calling participation practices on participation credit, uncapped magnitude of participation, participation rate, attendance, and adjusted exam scores. These dependent measures were compared between (a) voluntary and cold-calling conditions and (b) high and low participants under baseline (voluntary participation without credit and high-rate and low-rate participants). The use of voluntary and cold-calling procedures was alternated across units. Results were evaluated using mixed designs with repeated-measures across treatment units and between-subject comparisons.

For both capped and raw participation, students exhibited higher levels of participation during voluntary units. Students who were high in baseline raw participation remained significantly higher than the low group in raw participation earned. Raw participation of the high group was significantly higher during voluntary units; however, the low group did not differ significantly between voluntary and cold-calling. Overall, participation rate did not differ significantly between voluntary and cold-calling units. The low-rate group generally had higher participation rates under the cold-calling condition, whereas the high-rate group had greater participation rates under the voluntary condition. Attendance did not differ significantly between voluntary and cold-calling units. While students in the late onset condition did not differ in exam performance, students in the early onset condition scored significantly lower on exams during cold-calling units than during baseline. For exam performance, the main effect for treatment condition was not significant. A student survey revealed that a majority of students favored a voluntary participation arrangement. A majority of the students reported feeling

nervous during cold-calling units, but indicated they followed the discussion more closely during those units.

Advantages and disadvantages can be identified for both cold-calling and voluntary participation. Initially reticent students will likely become more engaged in class discussion under the cold-calling condition, whereas participation for the whole class will be higher under the voluntary condition. Some blending of the two conditions would probably be optimal: starting a class with cold-calling and gradually switching to voluntary participation as student engagement increases.

Table of Contents

List of Tables

List of Figures

Chapter I

Introduction and Literature Review

Instructors must identify critical strategies for engaging undergraduate students in classroom learning activities. Encouraging students to participate in class discussion is one engagement strategy that has been linked to a variety of achievement measures, such as increased critical thinking and problem-solving skills (Garside, 1996; Murray & Lang, 1997). If an instructor's primary goal is to optimize student engagement in class discussion, determining whether to call on students or solicit voluntary participation may be a primary consideration that affects quantity, and balance of participation across students, as well as major performance measures.

Role of Discussion in Course Engagement

Several researchers have included participation in class discussion in their overall scheme of student engagement in class. Dancer and Kamvounias (2005) included preparation and attendance in addition to contribution to class discussion and communication skills in their five-component model of course engagement. Likewise, Fritschner (2000) included attending class, notetaking, completing assignments, asking questions in class, making comments, providing input for class discussions, and conducting outside research as forms of course engagement in her six-component model. More specifically, Rocca (2010) defined class participation as consisting of asking questions, raising one's hand, and making comments when not requested by an instructor. Participation has also been defined as "the number of unsolicited responses volunteered by individuals" (Burchfield & Sappington, 1999, p. 290). Although some researchers (Dancer & Kamvounias, 2005; Fritschner, 2000; Petress, 2006) include nonverbal behaviors (e.g., attendance, raised hands, head nods, visual orientation) in their description of

participation in class discussion, others (Burchfield & Sappington, 1999) focus exclusively on verbal contributions in the class.

Relationship of Class Participation to Student Outcomes

Classroom participation may contribute to the ability to communicate effectively in group situations, a skill that has enduring and pervasive benefits in one's personal and professional development (Armstrong & Boud, 1983). Participating in class discussion has been shown to increase understanding of content knowledge and problem-solving skills (Murray & Lang, 1997). In the same manner, participation may foster critical thinking skills, particularly when discussion includes problem solving (Garside, 1996). Classroom discussion can also foster development of critical understanding, self-awareness, and an appreciation for diverse perspectives (Brookfield & Preskill, 2012). Additionally, participation has also been linked to increased motivation and self-reported gains in character development (Junn, 1994; Kuh & Umbach, 2004).

Measurement of Class Participation

Several factors must be considered when measuring classroom participation, such as how participation will be recorded and who will record it. Petress (2006) stated that participation must be evaluated either as a frequency count of student comments or as a measure of individual student progress in quality of participation. Recent research has indicated that measuring the quality of students' participation is equally predictive of course performance as measuring the quantity of student participation (Carstens, Wright, Coles, McCleary, & Williams, 2013). Other researchers have listed several reasons not to include participation in grading criteria: the lack of instruction on how to improve participation, the potential for instructor subjectivity, the dependence of participation on a student's personality or willingness to participate, and demands

of record-keeping (Jacobs & Chase, 1992). Also, the student's sex may confound the measurement of participation, as both male and female teachers have been found to devote more attention to male students in class discussion (Spender, 1982).

Researchers have used various methods to measure class participation. Early on, observation systems differentiating instructor talk, student talk, and silence were used to evaluate verbal activity within the classroom (Flanders, 1962). Karp and Yoels (1976) used an observation system that required the presence of a researcher. This observation method, however, may not discriminate between types of contributions (quantity versus quality) and may be time-intensive. Requiring instructors to record daily participation may interfere with classroom flow and chemistry of the class; however, waiting until the end of a course or even the end of a class period to record discussion may cause undue reliance on student and/or instructor memory (Armstrong & Boud, 1983).

External observers may be beneficial in preventing students from reporting inflated levels of participation, thereby preventing instructor biases from distorting the evaluation of participation (Armstrong & Boud, 1983). A feasible arrangement for computing inter-rater agreement between student and observer ratings would consist of external-observer availability at least on an intermittent basis. Additionally, students have also monitored and recorded their own participation, though reliability of student records has been mixed (Burchfield & Sappington, 1999; Carstens, Wright, Coles, McCleary, & Williams, 2013; Dancer & Kamvounias, 2005; Krohn et al., 2011; Melvin, 1988).

Contributors to Participation in Class Discussion

Although participation has generated numerous benefits, many students choose not to participate in class; historical reports indicate that only a handful of students participate during

class discussion (Karp & Yoels, 1976; Rocca, 2010). Specifically, more recent research has indicated that only 12% of students participated regularly and 25% participated intermittently (Weaver & Qi, 2005). In addition to finding that less than half of the observed students participated, Howard and Henney (1998) found that approximately 92% of all interactions were made by a small group consisting of about five students.

Both class size and seating arrangement have repeatedly been shown to affect student participation, with larger classes inhibiting student participation more than smaller classes (Constantinople, Cornelius, & Gray, 1988; Fassinger, 1995; Howard & Henney, 1998; Karp and Yoels, 1976). Larger classes provided more opportunity for anonymity and less opportunity to participate in discussion, whereas smaller classes limited the possibility of student withdrawal from active involvement (Weaver & Qi, 2005). The seating arrangement of a classroom has also been shown to affect student participation levels (Brown & Pruis, 1958; McCorskey & McVetta, 1978; Morrison & Thomas, 1975). Rocca (2010) suggested that U-shaped, circular, or semicircular seating arrangements allowed for greater participation than traditional row and column seating. Of course, size of a class will affect the types of seating arrangements logistically feasible.

In addition to class size and seating arrangements, instructor cues also influence class participation. For example, wait time after an instructor poses questions may affect levels of class participation. Students have interpreted minimal wait time as an indication that instructors did not desire participation (Fritschner, 2000). Perhaps for that reason, Bean and Peterson (1998) recommended increasing wait time after instructors pose questions. Furthermore, adequate student preparation for class discussions can facilitate participation (Fassinger, 1995). A lack of preparation may increase fears of peer or instructor disapproval of an inaccurate

comment (Weaver & Qi, 2005). Lack of preparation may also contribute to the need for extended wait time following instructor solicitation of student comments.

Grades may serve as one of the greatest incentives for class participation. Receiving credit for participating during class has been shown to increase students' overall contributions to the discussion, sometimes by as many as eight comments per class (Boniecki & Moore, 2003; Sommer & Sommer, 2007). Boniecki and Moore found that providing extra credit increased the number of hands raised, decreased the amount of wait time following an instructor's question, and increased the number of questions or comments from students. Having students evaluate their own participation throughout the grading process has been shown to increase both the frequency and the quality of participation (Zaremba & Dunn, 2004).

Personality characteristics have also predicted levels of student participation (Armstrong & Boud, 1983; Galyon, Blondin, Yaw, Nalls, & Williams, 2012; Morrison & Thomas, 1975; Weaver & Qi, 2005). Anxiety and tension may inhibit participation due to the threat of "appearing stupid in front of peers and teacher" (Armstrong & Boud, p. 37). A student's level of self-efficacy has predicted student participation and exam performance (Galyon et al., 2012). Students with low self-esteem spoke less and shared a smaller portion of their thoughts than students with high self-esteem; the former were also more likely to sit in the back of a classroom (Morrison & Thomas, 1975). Moreover, student assertiveness has been implicated in determining whether a student will participate in discussion (Rocca, 2010). Confidence is another trait commonly associated with participation. Weaver and Qi suggested that confidence serves to generate energy within a classroom, which leads to greater participation within a group of students. Students reported confidence to be the most important factor affecting levels of

participation, whereas higher levels of insecurity produces lower levels of participation (Fassinger, 1995; Williams, 1971).

Cold-calling Participation

One of the potentially greatest contributors to the nature, quality, quantity, and distribution of participation may be among the least researched—allowing students to volunteer their comments versus randomly calling on students to respond to instructor questions (Bean & Peterson, 1998). Some instructors consider the latter (cold-calling) to be punitive, humiliating, and cold, possibly causing students to feel uncomfortable or victimized and less likely to participate in the future. On the other hand, other instructors consider cold-calling to facilitate class discussions by maximizing student preparation for discussion, including more students in the discussion, and ensuring learning objectives are met. The "cold-calling" professor has been characterized as one who seeks "quality of response during the Socratic examination" (Bean & Peterson). However, Jones (2008) questioned the similarity to the Socratic Method, claiming that often the primary motivation behind cold-calling is holding students accountable for reading assignments. Despite its possible benefits, cold-calling is not a common practice within classrooms, reportedly occurring in only about 10% of classrooms (Karp & Yoels, 1976).

Some research has used cold-calling to identify faculty behaviors or characteristics that enhanced the quality of participation and the effectiveness of the class discussion (Dallimore, Hertenstein, & Platt, 2004). However, these authors failed to note the frequency with which cold-calling occurred throughout the study, stating only that the instructor was experienced and effective in class discussion, primarily devoted class time to discussion, and regularly cold-called on students. A survey asked students to indicate professorial behaviors that increased student participation and that either increased or decreased the effectiveness of the class discussion. Six

categories emerged from the survey responses: (1) requiring and grading participation, (2) incorporating instructor and students' ideas and experiences, (3) actively facilitating discussion, (4) asking effective questions, (5) creating a supportive classroom environment, and (6) affirming student contributions and providing constructive feedback. Cold-calling was considered to be a component of the first category—requiring and grading participation. All six of these categories, including the cold-calling embedded in component number one, were identified by students as positively affecting participation and their comfort within the classroom.

In later research, Dallimore, Hertenstein, and Platt (2006) operationally defined cold-calling as "any instance in which a teacher calls on a student whose hand is not raised" (p.355). The exploratory study consisted of a one-group, pre-post design, which limited the generalizability of its findings. Students were told they would be called on even if they had not raised their hand, with participation counting for 40% of a student's final grade. The frequency of cold-calling was not recorded; instead, the authors simply noted that the instructor used cold-calling extensively. This assertion was based on faculty members' past observations of the instructor's discussion style. Analyses consisted of a quantitative description of responses to a pre- and posttest questionnaires and a path analysis. The path analysis indicated that cold-calling and graded participation may increase participation frequency. Furthermore, the questionnaire analyses revealed that cold-calling did not make students uncomfortable in class.

Suggestions for increasing student comfort with cold-calling include providing response preparation time during class, making questions appropriately difficult, and calling on a variety of students (Dallimore et al., 2006). These authors suggested giving students advance notice that they may be called upon (e.g., telling a student he or she may be called upon prior to class and

identifying the question the student might be asked; posing the question to the entire class and allowing them time to reflect upon the question prior to cold-calling; and providing opportunities to discuss questions in small groups prior to cold-calling on specific individuals). The researchers claimed that using simpler questions early in the course may build confidence and better prepare students for cold-calling later in the course. The Dallimore et al. study, however, failed to provide both a comparison group and a measure of cold-calling.

A recent study by Dallimore, Hertenstein, and Platt (2013) evaluated the effects of cold-calling in several sections of an undergraduate class. Students were administered a pre- and post-course questionnaire and observed twice during the course. Observations were conducted on class discussion days by graduate research assistants, with only one observer present per observation day. Observers recorded which student responded to a question and whether the student volunteered a response or was called upon. The frequency of cold-calling was not regulated; in fact, instructors were not given any instruction as to how they should conduct their classes. Classes were categorized as either high or low cold-calling, with categorization based on the overall mean percentage of students cold-called within the class. High cold-call classes had a mean percentage of cold-called students ranging from 33 to 84%, while low cold-called classes had a mean percentage of cold-called students ranging from 0 to 24%.

Data analyses in the Dallimore et al. (2013) study revealed that the mean percentage of students who voluntarily answered questions was higher in high cold-calling sections than in low cold-calling sections. From the first to second observation, voluntary participation increased significantly within the high cold-calling classes but remained stable within the low cold-calling classes. Overall, frequency of participation (the number of questions answered per student) also increased significantly across observations in the high cold-calling sections. Likewise, overall

frequency was significantly higher in the high cold-calling classes than in the low cold-calling classes. Questionnaire analyses revealed no differences in reported levels of classroom comfort between students in high or low cold-calling classes.

Findings from the most recent Dallimore et al. (2013) study have limited generalizability, given that all participants were enrolled in the same course. However, their study did include several different sections of that course. The researchers' failure to assess inter-rater agreement and their reliance on survey data also limited the generalization of their findings. The greatest limitation of their research may be its failure to manipulate comparisons between voluntary and cold-calling conditions.

Framework for the Current Study

The current study addressed the major limitations of previous research, and sought to extend the available literature on cold-calling vs. voluntary class participation. Previous research showed cold-calling to be positively related to student participation and comfort within the classroom. However, these previous studies failed to provide adequate regulation of cold-calling, and classroom observation of participation was limited. Manipulating cold-called and volunteered participation would allow for much tighter cause and effect inferences regarding the impact of cold-calling versus volunteering on various participation and performance variables. Numerous observations of class participation and several inter-rater agreement checks would likely contribute to the reliability of the assessment procedures. Participation could then be compared across multiple sections of a course and multiple units within sections by utilizing an alternating schedule of both voluntary and cold-calling across units.

The number of studies on cold-calling versus volunteering is limited, and the few available studies lack rigorous assessment of treatment effects. Previous studies have neglected

to manipulate the cold-calling versus volunteering conditions or provide repeated and reliable measurement of class participation. Finally, none of the previous research reliably assessed the level of participation and precisely determined the balance of participation across students under the two arrangements. Plus, the effects of cold-calling versus volunteered comments on exam performance have not been evaluated.

Research Questions

The research questions addressed the effects of cold-calling and voluntary participation on various student outcomes and behaviors. One of the goals of the current study was to determine whether the two treatment conditions would differentially affect the amount and rate of participation within the total sample. An ancillary goal was to determine the effects of voluntary versus cold-calling on the participation of students identified in the baseline period as high or low participants. The dependent measures in all treatment comparisons were capped participation levels, raw participation totals, and rate of participation (individual number of comments divided by total comments in the class). Other goals were to determine whether the treatment conditions would differentially affect attendance and exam performance. The current study also sought student opinions about cold-calling vs. voluntary conditions and their own behaviors during the different participation conditions.

Chapter II

General Methods

Participants

Whole-sample participants. The study was conducted in eight sections of an undergraduate educational psychology course at a large Southeastern university. The course is required for those entering the University's Teacher-Education program. Due to the reduced class size (10-22 students in each section) and differences in treatment implementation, two of the sections were excluded from the final analyses. Thus, the final sample included six sections of the course. Most sections were comprised of approximately 25-30 students (N = 156). Participants were predominately female (85.9%). Participants ranged from freshmen through graduate students. However, the majority of students were sophomores (54.5%), with juniors comprising the second largest group (29.5%). Participants reported an average course load of 15 hours for the semester and an average work load of 12 hours each week. Students' mean reported grade point average was 3.37.

Levels of participation. To assess balance of participation across class members, students were divided by quartile ranks according to their raw participation totals during Unit A. The top and bottom quartiles of students' raw participation during baseline were used in subsequently comparing the effects of the treatments on the dependent variables in the remaining units of the course. The goal was to determine if cold-calling increased levels of participation for students with initially low levels of participation and decreased levels of participation by those initially inclined to dominate classroom discussions.

Students were also categorized using quartile ranks as high and low rate participants based on their rate of participation during Unit A. This distinction between high and low rate

participants was used to compare changes in the rate of participation across treatment conditions. The objective was to evaluate whether cold-calling increased the rate of participation for students less likely to contribute to the discussion. The percentage of the class in both the high raw participation and high rate groups was 17.3%. The percentage of students in both the low rate and low raw participation groups was 22.4%.

Course Structure

The course structure was consistent across the six sections. All sections were divided into five units (Units A-E) reflecting various human-development themes: physical, cognitive, values, social, and psychological development. All sections of the course used the same unit schedule, course materials, unit exams, and instructional approach. The grading structure of the course can be found in Appendix A. Students were asked to prepare for four in-class discussion days in each unit by reviewing instructor notes and answering questions in writing pertaining to those notes prior to their discussion in class. The predominant pedagogy for the class consisted of instructor-led discussion addressing questions students answered prior to class. Graduate teaching assistants (GTAs) served as the primary instructors in each section. The GTAs were trained to lead the class discussion by posing questions similar to those answered by students over instructor notes prior to class. Informal implementation integrity data was collected during inter-rater days by the experimenter. The total number of deviations made by the instructor from the intervention plan was recorded and totaled across all five observations. Weekly meetings were held to provide supplemental training if needed, as well as address any implementation concerns.

All sections followed the same approximate schedule for each of the five units: the first day consisted of viewing a video viewing and/or discussion of the video as related to the current

material and also included a structured discussion covering a portion of the prepared questions; the next three days were for structured discussion of the remaining discussion questions and reviewing a practice examination taken by the students outside of class; the final day was devoted to a 50-item multiple-choice exam that covered all information related to that unit (including instructor notes, PowerPoint slides, video, and journal articles in the unit).

Students submitted a record card at the conclusion of each discussion day (four days per unit). Record cards provided spaces for students to record their attendance, display of name card, number of instructor-notes questions answered, number of video questions answered, and number of article questions answered. Students received 2 points of credit for attendance each day (totaling 8 points across each unit) and 1 point of credit for the presence of their name card. The cards also had space for students to record their comments during the discussion, as well as their qualitative rating of each comment. Students were instructed at the beginning of the course to use instructor feedback as cues for rating their comments. Qualitative categories included the following instructor ratings: 0-point comments—redundant, off-topic, or totally incorrect; 1-point comments—partly correct; and 2-point comment—entirely correct or informed. The qualitative ratings were used in determining the amount of credit students received for participation. Students recorded volunteered comments on the front of the card and cold-called comments on the back of the card.

Dependent Measures

Several variables were measured across participants, including attendance, amount of participation credit earned by students, total amount of participation irrespective of credit, rate of participation, distribution of participation across students, and exam performance. Student's turned in record cards at the conclusion of each class period; these cards provided the

information for number of comments made by students and the qualitative rating of those comments. The student record card had space for recording three voluntary comments on the front of the cards and three cold-called comments on the back of the card; cold-called comments were referred to as "called-on" comments on the record card. If a student exceeded more than three voluntary or called on comments in a class session, he/she was instructed to record the fourth comment at the bottom of the appropriate side of the record card.

On one discussion day in each course unit, two observers recorded and rated each student comment. One of the observers in each section was a non-teaching graduate teaching assistant and the other observer was the experimenter. The observers sat in a front corner of the room where they could see each student's name card. Both the observers and the students were instructed to consider instructor feedback in rating each comment. Observers used a different form for each unit that listed each student's name, as well as identifying information regarding class sections and units (see Appendix B). They recorded the qualitative rating of each comment or question under the unit column, which was labeled with the treatment condition. Although some units in all sections involved primarily voluntary or cold-called comments, the order in which voluntary and cold-calling was used differed across sections. Thus, the record card had space for students to record both voluntary and cold-called comments each day. For example, if a student made a voluntary comment during a cold-calling unit, the observer recorded the comment under the "Voluntary" column in the appropriate unit column.

The participation credit earned by each student was recorded daily on student record cards; students were expected to accurately record a summary and qualitative rating of each comment made. Their qualitative rating (level 0, 1, or 2) specified the participation credit awarded to each student. Because quantity of participation has been shown to be equally

predictive as quality of comments, the qualitative ratings were used only as a measure of the amount of credit students earned. The mean credit earned across each unit was used in determining participation differences across treatment onset and discussion conditions. For grading purposes, a cap (6 points daily, 20 points per unit) was placed on the maximum number of points a student could earn from classroom participation across each unit. Students were made aware of the cap placed on participation credit; however, some students exceeded this maximum credit limit during both voluntary and cold-calling units.

Across the semester, students were given five multiple-choice exams, each composed of 50 questions. The exams were designed to require a critical evaluation of issues addressed in the course materials. Historically, exam means from 2004 through 2011 have been as follows: Unit A 39.42 (78.84% of possible credit), Unit B 36.68 (73.36%), Unit C 40.51 (81.02%), Unit D 38.91 (77.82%), and Unit E 39.24 (78.48%) (Galyon, 2012). Because course records over the last seven years revealed that exam scores are typically higher or lower in certain units, the current raw exam scores were converted to z-scores to account for historical differences in exam scores across units. The z-score for each student was computed by determining the difference between the historical mean and an individual's current exam score for a particular unit and then dividing that difference by the standard deviation of the historical scores.

Treatment Conditions

The study examined the effects of two treatment conditions on the specified dependent variables: (1) voluntary versus cold-calling participation in class discussion, and (2) point in the course when cold-calling was implemented (earlier or later). The first unit of the course (Unit A) was devoted entirely to collection of baseline data on voluntary participation with no credit granted for participation. This period was used for students to practice recording and rating their

comments using the record card. Instructors were explicit in providing feedback during this time and provided examples of the qualitative ratings for student comments. External observers also practiced the use of their observational system in baseline. The experimenter provided corrective feedback to second observer when needed.

Participation credit was awarded for the subsequent four units (Units B, C, D, and E). Students received up to six points of participation credit each discussion day (four days per unit) and up to twenty participation points for each unit totaling eighty points for the course as a whole (equaling approximately 15% of a student's total grade). In three of the six course sections, instructors used cold-calling during Units B and D only and voluntary participation during Units C and E. The other three sections used voluntary participation during Units B and D, and cold-calling only during Units C and E. Instructions were posted on the course website and sent by email to students at the start of each voluntary or cold-calling unit (see Appendix C). Three GTAs taught two sections each, with one section using the first sequence of voluntary and cold-calling conditions and the second section using the opposite sequence. Table 1 shows the sequence of treatment across units in the six sections. This schedule was intended to determine whether the point of introducing cold-calling differentially affected attendance, frequency of participation, quality of comments, distribution of participation across students, and/or exam scores.

Voluntary participation units consisted of instructors posing a question or comment and asking for volunteers to respond by raising their hand. Questions asked by the instructor were based on questions given to the students prior to class and followed the order provided to students. Students were not called upon during voluntary units. If no student responded to a question or comment after 15 seconds, instructors rephrased the question or comment and again

solicited responses from other students. If no answer was provided, instructors moved on to the next question. Students were instructed at the beginning of voluntary units that they should participate freely. They were reminded of the participation component of their final grade and encouraged to voice their understanding of the course material and other students' comments regarding the course materials.

At the beginning of cold-calling units, students were informed that instructors would pose questions and then call on a student to provide an answer; students were not informed as to the order students would be called upon. Instructors announced that students were welcome to volunteer questions but that the instructor would otherwise randomly call on students two to three times in each class session. Should a student attempt to volunteer a comment, the instructor was to remind the student that comments should only be made when called on or to rephrase the comment into a question. After all students had been called upon during cold-calling units, instructors once more randomly called upon students but in a different random order.

To simplify the burden on instructors, the experimenter printed a randomized class roster prior to each class during cold-calling units; student names were listed three times in three different random orders. Instructors made a mark next to each student's name when that student was called on, and then moved to the following name. Although students were permitted to volunteer questions throughout the called-upon units, virtually all instructor questions were followed by the instructor's calling on a specific student to answer that question. Occasionally, if an instructor failed to call on a specific student during a given class period, the instructor compensated for that omission the following class period by calling on that student the requisite additional times. Because not all students were called upon during a given class period, a

participation rate was calculated for each student by dividing the total number of comments made by each student by the total number of comments made during each class session.

At the conclusion of the semester, students were asked to complete a survey regarding their perceptions and opinions of the two treatment conditions. The survey consisted of 25 questions to be completed outside of class and answered on a scantron. Survey responses were based on a 5-point Likert Scale with possible responses including strongly agree, agree, neutral, disagree, or strongly disagree. Participation in the survey was voluntary and students were not given course credit for completing it.

Chapter III

Results

To determine differences in treatment effects of voluntary and cold-calling conditions, I used a mixed-factor design with baseline and the treatment units as repeated measures and earlier vs. later introduction of cold-calling as the between-subjects measure. In the second phase of the analyses, the top- and bottom-quartile voluntary participants during baseline were compared on the participation measures across the subsequent treatment units. Participation levels across units served as the repeated measure and participation levels for the high and low baseline participants as the between measure. This analysis permitted assessment of the main effects of treatment units and the two participation groups, as well as interaction between these two independent variables.

Implementation Integrity

Implementation integrity data was collected during inter-rater days by the experimenter. The total number of deviations made by the instructor from the intervention plan was recorded and totaled across all five observations. Each section was then categorized as having poor, medium, or strong integrity based on these totals. Sections 7 and 8 of the original sample, both taught by the same instructor, had higher levels of integrity the other 6 sections. Because of this and class size, sections 7 and 8 were excluded from subsequent analyses. These sections likely had higher integrity due to the reduced class size, which led to students being called on and having opportunities to volunteer much more frequently than students in the other six sections.

Inter-rater Agreement

For one discussion day in each unit in each section, observers recorded and rated each student comment. A percent inter-rater agreement score was computed by finding the percentage

of agreement between observer and student records (Carstens et al., 2013). For example, if a student reported earning 5 points and an observer reported the student's earning 4 points, the percentage of agreement would be 80%. Agreement scores were computed for amount of participation between student and observer records and also between the two observers. Percentage of agreement ranged from 69-100% (see Table 2). Observers had much higher levels of agreement with one another than with students; observer agreement ranged from 91-100%.

Capped Participation Credit

Students' mean capped credit earned across treatment units was 15.80, 14.71, 14.65, and 14.21 for Units B, C, D, and E, respectively. If the same credit criteria had been applied to Unit A participation, the credit earned would have been 13.99. The mean capped credit earned was 15.90 for the voluntary units and 14.78 for the cold-calling units for Units B-E (see Table 3). The mean capped credit for the voluntary units was 15.42, 15.92, 15.99, and 16.12 for Units B, C, D, and E, respectively. For cold-calling units the capped credit means were 16.17, 13.50, 13.30, and 12.30 for Units B, C, D, and E, respectively.

Capped participation during Unit A, voluntary capped participation earned, and cold-calling capped participation earned served as the within-subjects variable, while the treatment onset condition (early v. late onset of cold-calling) was a between-subjects variable. A repeated measures ANOVA did not yield a significant interaction effect for onset condition by participation condition, $F(2, 308) = 2.32$, *ns*. A significant main effect was obtained for participation condition, with voluntary means, cold-calling means, and Unit A means differing significantly, $F(2, 308) = 10.40$, $p < 0.001$, partial eta squared = 0.63, power = 0.99. Students earned more capped credit during the voluntary units ($M = 15.90$, $SD = 4.96$) than the cold-

calling units (M = 14.73, SD =3.52) and Unit A (M = 14.06, SD =5.80). However, Unit A and cold-calling did not differ significantly in the amount of capped participation.

Raw Participation

Whole-sample differences. The uncapped, raw amount of participation across units was 15.79, 17.04, 15.88, 15.72, and 15.31 for units A, B, C, D, and E, respectively. The mean raw amount of participation was 17.52 for the voluntary units and 14.46 for the cold-calling units (see Table 4). A repeated measures mixed-factor design examined the effects of the within-subjects participation arrangements (voluntary, cold-calling, and baseline) and between-subjects onset condition (earlier vs. later introduction of cold-calling) on raw participation levels. This analysis yielded no significant interaction between the onset and participation conditions, $F(2, 308) = 1.09$, *ns*. However, a significant main effect was obtained for raw participation levels, $F(2, 308) = 17.17$, $p < 0.001$, partial eta squared = 0.10, power = 1.00. Students had higher levels of participation during voluntary units (M = 17.52, SD = 6.32) than cold-calling units (M = 14.46, SD = 4.08) and Unit A (M = 15.79, SD = 7.66). Unit A and cold-calling raw means did not differ significantly.

Group levels of participation. The high-participant raw-participation means were 24.85, 18.98, 18.61, 18.34, and 18.27 for Units A, B, C, D, and E, respectively. The corresponding means for low participants were 6.00, 14.18, 11.98, 12.10, and 11.10 for Units A, B, C, D, and E, respectively. High participants had a mean participation level of 20.84 during voluntary units as compared to 12.05 by low participants. During cold-calling units, high participants had an average of 16.26 and low participants 12.63 for participation magnitude (see Table 5).

A mixed-factor ANOVA used voluntary and cold-calling conditions as the within-

subjects variable and initial participation level (high v. low) as the between-subjects variable in examining effects of the two treatment conditions on the raw participation of high and low participants. A significant interaction effect was revealed between treatment condition and participation level on raw participation totals, $F(1, 79) = 20.03$, $p < 0.001$, partial eta squared = 0.20, power = 0.99 (see Figure 1). A simple-effects analysis showed that high participants had significantly higher levels of participation in voluntary units ($M = 20.84$, $SD = 4.69$) than cold-calling units ($M = 16.26$, $SD = 3.46$), $F(1, 40) = 50.64$, $p < 0.001$, partial eta squared = 0.56, power = 1.00. In contrast, low participants had no significant raw-participation differences between the voluntary and cold-calling units (see Table 6).

The analysis revealed a significant main effect for group participation level, with high participants ($M = 18.55$, $SD = 4.07$) showing higher levels of participation across units than low participants ($M = 12.34$; $SD = 5.32$), $F(1, 79) = 45.47$, $p < 0.001$, partial eta squared = 0.37, power = 1.00. The main effect for group raw participation was significant, $F(1, 79) = 12.10$, $p < 0.01$, partial eta squared = 0.13, power = 0.93. Students had higher levels of participation during voluntary units (M = 16.50, SD = 7.37) than cold-calling units (M = 14.46, SD = 3.98). Figures 2 and 3 show the pattern of participation means across the late and early onset conditions for high and low participants.

Rate of Participation

Whole-sample participation rates. The overall numbers of opportunities to participate are recorded in Table 7. The average number of opportunities to participate was higher in voluntary units (250) than in cold-calling units (220). Participation opportunities were 242, 261, 234, 226, and 218 for Units A, B, C, D, and E, respectively. For the voluntary condition, the mean participation opportunities were 246.33, 252.33, 251.00, and 250.00 for Units B, C, D, and

E, respectively. The mean participation opportunities for the cold-calling condition were 275.33, 215.00, 201.00, and 186.67 for Units B, C, D, and E, respectively. Thus, with the exception of Unit B, opportunities to participate were less under cold-calling than voluntary participation.

Chi-square Goodness of Fit tests were conducted to determine differences between voluntary and cold-calling opportunities in Units B-D and the total number of opportunities across units. In Unit B, there were significant differences in actual opportunities to participate between voluntary and cold-calling conditions compared to the expected number of opportunities, with cold-calling units having a greater number of actual opportunities, $\chi 2$ (1, n = 1565) = 4.84, $p < .05$. Though the difference was also significant for Unit C, voluntary units produced a greater number of opportunities than cold-calling units, $\chi 2$ (1, n = 1402) = 8.95, $p <$ .05. Unit D followed the same pattern with voluntary units having a greater number of opportunities, $\chi 2$ (1, n = 1356) = 16.59, $p < .05$. Unit E had significant differences between voluntary and cold-calling conditions as well, $\chi 2$ (1, n = 1310) = 27.56, $p < .05$. Voluntary units had a higher number of opportunities to participate during Unit E. Across units, there were significant differences between the total number of voluntary opportunities and the number of cold-calling opportunities compared to expected opportunities under those conditions, $\chi 2$ (1, n = 5633) = 23.65, $p < .05$. It was expected that a similar number of opportunities would be available during both voluntary and cold-calling units; however, chi-square analyses revealed significantly higher number of opportunities under the voluntary condition.

Because the number of opportunities differed significantly, participation rates were calculated for each student by dividing the total number of comments made by an individual student by the total number of comments made during the class period. For example, if a student made 5 comments during a class period in which a total of 225 comments were made, the student

would have a participation rate of 0.02 or 2%. Mean rates of participation were 3.55%, 3.59%, 3.59%, 3.60%, and 3.65% for Units A, B, C, D, and E, respectively. The rates for both voluntary and cold-calling units were similar at 3.60% and 3.62%, respectively.

A repeated measures ANOVA was run using participation rate during Unit A, first voluntary unit rate, second voluntary unit rate, first cold-calling rate, second cold-calling rate as within-variables. Onset condition was used as a between-subjects variable. No interaction or main effects were significant for participation rates. Another repeated measures ANOVA was run using voluntary rate during Unit A, voluntary units, and cold-calling as within-variables, and onset condition was used as a between-subjects variable. No interaction or main effects were significant. Thus, participation rates were equivalent under voluntary and cold-calling conditions or under earlier v. later onset of cold-calling.

Group-level participation rates. Based on participation rates during baseline, top and bottom quartile students were categorized as high- or low-rate participants. The mean participation rates for high participants were 5.49%, 4.16%, 4.20%, 4.29%, and 4.27% for Units A, B, C, D, and E, respectively. Low-rate participants had mean rates of 1.46%, 2.93%, 2.82%, 3.03%, and 2.70% for Units A, B, C, D, and E, respectively. High-rate participants earned an average rate of 4.51% during voluntary units as compared to the 2.44% rate earned by low-rate participants. During cold-calling units, high participants had an average rate of 3.95% and low rate participants earned 3.29% (see Table 8).

A repeated measures ANOVA compared participation rate during Unit A, first voluntary unit, second voluntary unit, first cold-calling unit, and second cold-calling unit as the within-subjects variable, and differential rate groups as a between-subjects variable. The interaction between participation group and discussion condition was significant, $F(3.41, 259.44) = 44.66$, p

< 0.000, partial eta squared = 0.37, power = 1.00. Simple effects showed a significant difference between high- and low participants across all within-subjects variables (rate during Unit A, first and second voluntary and cold-calling rates), with high-rate participants having higher rates of participation. Looking across the within-subjects measures with Bonferroni pairwise comparisons, the experimenter found that low participants had a significantly lower rate in Unit A than in any of the other units. They had significantly higher rates during the first cold-calling unit than both voluntary units. The second cold-calling unit was significantly higher than the first voluntary unit, but not the second. The two voluntary units did not differ significantly from one another. High participants had significantly higher rates of participation in Unit A than across voluntary or cold-calling units. Rates were significantly lower in the first cold-calling units than in the first voluntary unit for high participants, but not different in the second cold-calling and voluntary units. A main effect for the high and low participants was significant, $F(1, 76) = 118.99$, $p < 0.000$, partial eta squared = 0.61, power = 1.00, with high-rate participants having higher rates of participation.

A repeated measures ANOVA was run with Unit A, voluntary units, and cold-calling units as the within-subjects variable and group participation-level as the between-subjects variable. An interaction effect between participation group and discussion condition was significant, $F(2, 152) = 76.67$, $p < 0.000$, partial eta squared = 0.50, power = 1.00 (see Figure 4). Simple effects showed a significant difference between high and low participants across all within-subjects variables, with high participants having higher rates of participation. A main effect for the high and low participants was also significant, $F(1, 76) = 185.21$, $p < 0.000$, partial eta squared = 0.71, power = 1.00, with high participants having higher rates of participation. Bonferroni pairwise comparisons across the within-subjects variables revealed that low

participants had a significantly lower rate in Unit A than the other conditions. Low participants had significantly higher rates during the cold-calling than the voluntary units. Again, high participants had significantly higher rates of participation in Unit A than across voluntary or cold-calling units. High participants had significantly higher rates during voluntary units than the cold-calling units.

Attendance

Unit A attendance, voluntary attendance, and cold-calling attendance served as a within-subjects variables, while the treatment onset (early v. late onset of cold-calling) was a between-subjects variable. A repeated measures ANOVA did not reveal a significant interaction effects between treatment onset and treatment condition for attendance, $F(2, 308) = 1.53$, *ns*. The main effect for attendance was significant, $F(2, 308) = 3.89$, $p < 0.05$, partial eta squared = 0.03, power = 0.70. Students showed higher attendance during Unit A ($M = 7.55$, $SD = 1.01$) than cold-calling units ($M = 7.31$, $SD = 0.99$). Attendance in voluntary units ($M = 7.38$, $SD = 0.92$) did not differ significantly from attendance in either cold-calling units or Unit A (see Table 9).

Exam Performance

Whole-sample exam performance. The exam scores of each section were recorded, as were the average exam scores of each student earned across both voluntary and cold-calling units (see Table 10). The overall mean exam scores were 41.40, 39.13, 43.13, 41.72, and 39.28 for Units A, B, C, D, and E, respectively. The mean exam scores were 40.86 for the voluntary units and 40.77 for the cold-calling units. Average z-scores were 0.36, 0.37, 0.62, 0.23, and 0.06 for Units A, B, C, D, and E, respectively (see Table 11).

Voluntary and cold-calling exam z-scores, as well as Unit A exam performance, served as the within-subjects variable, while the treatment onset condition (early v. late onset of cold-

calling) was a between-subjects variable. A repeated measures ANOVA yielded a significant interaction effect for onset condition and exam performance, $F(2, 308) = 4.47$, $p < 0.05$, partial eta squared = 0.03, power = 0.76 (see Table 12 and Figure 5). Simple effects analysis was run to determine the nature of the significant interaction. Students in the late onset did not differ significantly across Unit A performance, voluntary performance, or cold-calling performance, $F(2, 156) = 0.99$, *ns*. Students in the early onset condition differed significantly across discussion conditions, $F(2, 308152) = 4.07$, $p < 0.05$, partial eta squared = 0.05, power = 0.72. Students scored lower during cold-calling units ($M = 0.27$, $SD = 0.79$) than during Unit A ($M = 0.49$, $SD = 0.83$). There were no exam differences between voluntary units ($M = 0.32$, $SD = 0.74$) and Unit A or cold-calling units. The main effect for treatment condition was not significant, $F(2, 308) = 0.47$, *ns*. The main effect for onset condition was also non-significant, $F(1, 154) = 0.22$, *ns*.

High- and low-level participants. Differences between high and low participants in exam performance were evaluated using a repeated measures ANOVA. Exam z-scores in Unit A, the first voluntary unit, the second voluntary unit, the first cold-calling unit, and the second cold-calling unit constituted the within-subjects variable and level of group participation (high/low) as a between-subjects variable. The interaction between participation group and treatment condition for exam z-scores was not significant. The main effect of the discussion condition on exam z-scores was significant, $F(3.47, 263.94) = 6.02$, $p < 0.000$, partial eta squared = 0.07, power = 0.97. The second cold-calling unit exam z-score was significantly lower than all other within-subjects variables (Unit A and first and second voluntary and cold-calling exam z-scores) except the second voluntary unit z-score. The first voluntary unit z-score was also significantly higher than the second voluntary unit z-score. The main effect for participation

groups was also significant, with high participants having a higher exam z-score mean (high group: $M = 0.57$, $SD = 0.75$; low group: $M = 0.17$, $SD = 0.86$), $F(1, 76) = 7.98$, $p < 0.000$, partial eta squared = 0.27, power = 1.00.

Survey Measure

Students across all six sections received the same survey at the conclusion of the course. Students could choose strongly agree, agree, neutral, disagree, or strongly disagree to survey items. Appendix F shows the means for all survey items. Many of the students reported they did not prefer a traditional lecture format (49.7%), while some students had no preference (22.8%). Generally, a majority of students believed they earned more participation credit during voluntary units (62.8%) and thought the course should be managed on a strictly voluntary basis (63.5%). They also disagreed that they did better on exams during the cold-calling units (disagreed or strongly disagreed = 42.1%; neutral = 39.3%; agreed or strongly agreed = 18.6%). A large majority (73.8%) reported feeling nervous during cold-calling units, despite many of the students' claim they felt more prepared during cold-calling units (52.4%). Some students reported tracking the discussion more closely during cold-calling units (44.8%) and that cold-calling units increased the probability that everyone would participate (47.6%).

The differences between high and low participants were examined using independent samples t-tests. Low participants (44.7%; $M = 2.95$, SD = 1.52) disagreed more than high participants (21.2%; $M = 1.76$, SD = 1.00) on the claim, "I earned more participation credit in the voluntary participation units than in called-on units", $t(69) = 3.82$, $p < 0.000$. Low participants (68.4%; $M = 2.16$, $SD = 0.89$) agreed more than high participants (33.3%; $M = 3.09$, $SD = 1.21$) with the statement, "I prepared more for called-on participation units than voluntary units", t (69) = -3.74, $p < 0.000$.

Chapter IV

General Discussion

The current study sought to extend the research findings regarding the effects of cold-calling and voluntary commenting on various measures of participation, exam performance, and attendance levels. Previous research has indicated cold-calling may increase the frequency of participation and does not negatively impact classroom comfort for students. By employing a design using both voluntary and cold-calling conditions as repeated measure, the experimenter differentiated the effects of the two discussion conditions on various dependent variables. The study expands current research by providing a structured examination of the effects of cold-calling through the manipulation of its use within a classroom.

Capped Participation

The participation cap of 6 points daily and 20 points per unit was applied to students' participation for grading purposes. Across units, students tended to earn between 14 and 16 points regardless of unit difficulty or treatment condition. On the average, students did not earn all of their potential points, tending to stop participating after earning a majority of the 20 points. While voluntary units showed higher levels of capped credit, students only earned an average of about 1 point higher than in cold-calling units. The 1 point difference ultimately accounted for only 0.19% of a student's final grade. Also, during cold-calling units, students did not differ from baseline in their participation levels.

Raw Participation

Whole-sample differences. Because some students chose to exceed the cap placed on credit, raw participation totals may provide a more accurate depiction of participation patterns. Students' raw participation approximated a magnitude of 16 points per unit, except Unit B where

they accumulated about 17 points. Raw participation followed a similar model to that of capped participation, with voluntary units producing higher levels of participation than cold-calling units. The difference in participation levels was about 3 points for raw participation (about 0.56% of a student's grade), while it was only 1 point for capped participation. This pattern may indicate the daily 6 point cap placed on credit each day may have controlled for some of the differences in participation levels between the two discussion conditions.

High- v. low-participation groups. High-level participants tended to participate less as the semester progressed. However, this decrease was slight with high participants earning approximately 18 or 19 points regardless of treatment unit. During Unit A, though, high participants produced their highest level of participation points (approximately 25). This is surprising given that no credit was earned for a grade during Unit A for participation, but was given in subsequent units. High participants earned a greater amount of participation points (4 points difference) during voluntary units than in cold-calling units.

Low-level participants displayed a more expected level of participation during Unit A, accumulating only about 6 points worth of participation. Whether this low level of participation was based more on initial reticence to participate or the lack of graded credit is unclear. Low-level participants accumulated anywhere from 11 to 14 points on average across units. Low participants accumulated slightly more participation points during cold-calling units, about half a point more, than voluntary units.

The interaction between participant group and treatment conditions was significant, with high participants accumulating significantly higher levels of participation points in the voluntary units. Low participants had no significant differences between accumulated participation during voluntary and cold-calling units. This finding indicated cold-calling may negatively impact

participation of initially high-level participants, but have no effect on participation of low participants. However, high participants still earned higher levels of participation across discussion conditions than low participants, and voluntary units produced higher levels of participation than cold-calling units for high participants. A convergence from baseline between the high and low participants emerged across the semester. High participants tended to decrease their participation in post-baseline units, while low participants increased their participation following baseline.

Rate of Participation

Total sample participation rates. Participation rate was calculated to lessen the impact of differences in opportunities to participate between voluntary and cold-calling conditions. The overall number of comments made decreased across the semester. More comments were made during baseline than in all subsequent units except one (Unit B). Again this is somewhat surprising, as students were not awarded grades for participating during Unit A. Voluntary units, on average, produced a higher number of comments than cold-calling units. Unit B was the only unit that produced a higher number of comments under cold-calling conditions. This was likely because Unit B material is considered more difficult and is often unfamiliar to the students. This increased difficulty likely inhibited students from volunteering more frequently due to the fear of answering incorrectly. Because generally a greater number of comments were being made during voluntary than cold-calling units, a rate variable was created to account for the discrepancy. While the number of comments made during each unit differed, the rate of comments made across the semester remained fairly constant in differing by only thousandths of a point. Similarly, voluntary and cold-calling produced the same rates of participation. Overall, the rate of participation was not affected by the discussion condition or the sequence of voluntary

and cold-calling units.

Group participation rates. Students were categorized according to their initial participation-quality points during Unit A into high- and low-level participants (top and bottom quartiles). High participants had an average rate of 0.05 during Unit A and averaged a rate of 0.04 across all other units. Thus, their participation rate diminished significantly following baseline. They exhibited higher rates of participation during voluntary units than cold-calling units, with about a 0.01 difference. Low-level participants, however, earned about 0.01 more during cold-calling units than voluntary units. Also, low-rate participants increased from a rate of 0.01 during Unit A to a 0.03 rate across the other units.

The interaction effect between participation rate levels and discussion condition was significant. Though high-level participants had higher rates than low-rate participants across discussion conditions, their patterns of participation tend to converge somewhat during the course of the semester (i.e., high-level participants were negatively impacted by cold-calling, while low-level participants were positively impacted). Participation rates were lowest during the first cold-calling unit but equivalent across the second cold-calling unit and the two voluntary units. This pattern may have been due to the change in the classroom discussion procedures; students may have initially been unsure about how to fully participate within the new cold-calling context.

Effects of Discussion Conditions on Attendance

Attendance was tracked for grading purposes as well as to determine if students would attend classes more or less frequently based on discussion condition. While there was a significant difference between baseline and cold-calling unit attendance, the mean difference was very slight (0.24 of a point) in favoring baseline. The lack of a difference in attendance across

voluntary and cold-calling units may indicate students were not uncomfortable during cold-calling units to the extent of attending class less frequently.

Effects of Discussion Conditions on Exam performance

Students scored between 39 and 43 mean points out of a possible 50 across all course units, equivalent to B and C letter grades. Mean exam scores were similar for both voluntary and cold-calling units. Because exam performance has a history of varying across units due to unit difficulty, z-scores were computed for each student's exam score to equate difficulty. While there were no significant main effects for discussion condition (voluntary or cold-calling) or onset of cold-calling (earlier introduction to cold-calling or later introduction), there was an interaction between these two variables. Students in the early onset condition, with cold-calling during Units B and D, scored significantly lower on exams during cold-calling units than during baseline. Students had experienced their initial exams under voluntary conditions and likely had an established method for class and exam preparation. The immediate introduction of cold-calling may have caused students to substantially change their preparation, which may have resulted in the lower exam scores. In the subsequent voluntary unit, students could revert to their primary method of preparation, which would lead to similar grades to those in baseline. In the second cold-calling unit, students would again have to alter their primary method of preparation. Students may have needed a longer time to acclimate to the course structure prior to changing the discussion format. In contrast, students in the late onset condition would have had additional time to master the course structure and their personal methods of exam preparation, which would have given them an advantage when the discussion format shifted to cold-calling.

High and low-level participant exam means were also evaluated. While there was no interaction between treatment condition and group participation level for exam z-scores across

voluntary and cold-calling units, there was a main effect for high v. low participants. High-level participants had higher exam means than low-level participants. Higher levels of participation may come from a greater understanding of course material, which would, in turn, produce higher exam scores. Additionally, exam scores varied across voluntary and cold-calling units, with the second cold-calling unit producing the lowest means in comparison to all other units except the second voluntary unit. This finding may provide some evidence to contradict the notion that cold-calling leads to greater preparation for exams. It may be that grades, in general, decreased as the semester continued; students may have begun to fatigue and consequently decrease their level of preparation for exams.

Participation Survey

The participation survey was given at the conclusion of the course to all students; it consisted of a 5-point Likert scale from strongly agree to strongly disagree. Students were asked several questions concerning preparation for class discussion, their perceptions of credit earned, and their overall comfort during voluntary and cold-calling units. Generally, students reported favoring the voluntary units, reporting the course discussion should be managed on a strictly voluntary basis. They also tended to believe they earned more credit during voluntary units, which may explain their preference for that arrangement. Students could have been dissatisfied with instructors' control over levels of participation during cold-calling, which may have had an adverse impact on their grades. Though a majority of students claimed they prepared more and a large number (44.8%) reported following the discussion more closely during cold-calling units, many (42.1%) students believed they did not do better on cold-calling exams. Their belief was accurate, as there were no differences in exam means across voluntary and cold-calling units. This finding leads to questioning whether students actually prepared more during cold-calling

units; additional preparation should lead to higher exam scores. However, the extra preparation and closer following of the class discussion may be a result of increased nervousness during cold-calling units. Almost three quarters of the students reported feeling nervous during cold-calling units.

High- and low-level participants differed on some survey items. Low-level participants were more likely to disagree that they earned more credit during voluntary units than cold-calling units. This claim was an accurate self-assessment, as low-level participants had higher rates of participation and earned slightly more credit during cold-calling than voluntary participation units. Low-level participants also tended to agree they prepared more for called-on units than voluntary units. This extra preparation may have led low-level participants to feel more comfortable during cold-calling units, which, in turn, led to higher rates and higher credit during those units.

Onset of Cold-calling Conditions

Because discussion conditions were alternated across units, some students were introduced to cold-calling earlier than others during the semester. Generally, whether students were in the late onset (cold-calling was introduced later in the semester) or early onset condition did not appear to have an effect on outcomes. There were no differences between early and late onset conditions for capped or raw participation credit, participation rates, or attendance. However, onset did have some effect on exam performance, with students in the early onset scoring lower during cold-calling units than during baseline.

Though there are limited effects of cold-calling onset on student outcomes, there may be some environmental benefits to introducing cold-calling earlier in the classroom. Survey results support the notions that cold-calling increases preparation and discussion engagement. Both

students and instructors may feel more comfortable with the procedure when introduced earlier, and students may attend more closely to the discussion. Students may also be encouraged to prepare for class discussions earlier in the semester. Cons may include intimidating students, frustrating high participants by limiting their contribution, and decreasing instructional time while keeping up with the cold-calling random regimentation.

Advantages and Disadvantages of Discussion Conditions

Voluntary participation. The voluntary participation condition typically produced higher levels of both capped and raw participation credit than the cold-calling condition. Voluntary participation was also preferred by students according to survey reports. However, during voluntary units, more reticent students were significantly less likely to have a higher rate of participation than during cold-calling units. There was also less balance in participation across students in the voluntary than cold-calling units.

Cold-calling participation. Cold-calling was shown to participation rate in initially reticent students. Alternatively, there was a decrease in participation for high-level participants. Cold-calling also produced a more balanced distribution of participation across students, though student reports of the condition were unfavorable. When participation frequency was divided by opportunities to participate, students' rate of participation proved equivalent under the two treatment conditions.

Limitations and Future Directions

A limitation of the current study was the difficulty instructors had in calling on each student an adequate and equivalent number of times. The number of comments made by each student during cold-calling units depended heavily on the instructor's ability to call on each student multiple times. The total number of comments made during cold-calling units was less

than the number made during voluntary units, indicating that students did not have as many opportunities to participate under cold-calling as under voluntary participation. The logistics required in implementing the cold-calling treatment slowed the pace of instructor's asking questions compared to the voluntary condition. It may also be that students were less likely to volunteer questions during cold-calling units or were unsure how to rephrase comments into questions that would elicit further discussion. Students may have felt the need to elaborate more extensively during cold-calling units due to social pressures, which would have decreased the amount of class time for other students to participate.

The discrepancy between the volume of voluntary and cold-calling comments made an accurate comparison of the two conditions difficult, which precipitated the use of the rate transformation. Future studies should seek to balance participation opportunities across cold-calling and voluntary conditions. A structure in which students are also able to volunteer comments during cold-calling could increase total comments made. It may also be necessary to train instructors to pace the discussion under cold-calling more efficiently to allow for maximum participation. Anecdotally, instructors reported difficulty calling on each student enough times to match opportunities to participate in the voluntary units. GTAs also reported difficulty with generating enough questions for each student to comment the necessary amount of times to provide opportunities to maximize their participation credit. In order for students to receive full credit for participation, they would need to comment 2-3 times each class period. In a class of 25 students, this required instructors to generate 50-75 questions or ask for multiple responses to the same question. Instructors struggled with continuing the discussion during cold-calling units more than during voluntary units. Students would frequently volunteer comments related to

another student's comment during voluntary units, but were unable to do so during cold-calling units. Instructors reported having difficulty bridging student comments to one another.

A second limitation was the use of self-recording practices for monitoring attendance and participation. Awarded participation credit was based on students' self-reported rating for each comment made. While it is possible for students to have inflated their participation credit, inter-rater agreement with observers was at an acceptable level to conclude that students did not inflate their scores. Additionally, the survey relied entirely on student opinion of their performance and behaviors during the semester. While students claimed to prepare more and follow the discussion more closely during cold-calling units, this was not evidenced in their grades. Thus, students may not be accurate in describing their own behaviors. A possible way to control for some of these self-report measures could be through the use of external observers. More observers present would allow more behaviors to be monitored. Behaviors such as physical orientation of a student and what content is on a student's computer would provide more information that may clarify findings.

Future research may wish to use external observers to track student participation, as well as instructor behavior during voluntary and cold-calling conditions. Though students indicated the instructor's were helpful with feedback and made the cold-calling units comfortable, student reports on the survey may not be accurate. Observing the instructors' behaviors and monitoring their implementation more closely would be the best way to ensure high treatment integrity.

A third limitation was the level of structure associated with the course used for the study. This course was highly structured, with students having prior knowledge of material to be covered in a given class period. This advance knowledge of the questions to be discussed during class allowed students to prepare for topics, which may have inflated participation. Under

voluntary conditions students may have prepared responses to specific questions and only participated when those topics were being discussed. During cold-calling conditions, students should have prepared responses and questions for all topics of the day; however, exam scores did not reflect a greater level of preparation during cold-calling units. Had students not known what questions would be asked of them, they might have been forced to prepare differently for class. Students were also aware of the discussion condition, which may have impacted their level of preparation and their participation during class. This high level of structure may limit the generalizability of findings. Future studies should examine the effects of limiting advance notice for students. Not giving students advance detailed knowledge of content to be covered or the discussion format to be used should change the way in which students prepare for class. Students might have prepared to a greater degree for cold-calling units than voluntary units.

Additionally, the course consisted of predominately female students who were in their sophomore and junior years of college. This limits the generalizability of the study's findings. It is possible that having a greater diversity in the students in a sample would affect the classroom discussion. Previous research indicated that instructors are more likely to favor males during class discussions (Spender, 1982). With a greater male presence, the balance of participation in the current study could shift. Also, advanced students (seniors and graduate students) may feel more comfortable participating in classroom discussion, which may alter the balance of participation as well as the quantity.

Conclusions

This study expands current research regarding the logistics and effect of cold-calling. Contrary to Dallimore et al. (2013) students participated less in the cold-calling units than the voluntary units, though this pattern was not true for those students who were initially lower

participants in the discussion. Cold-calling served to increase participation rates in students who were preliminarily reticent to contribute. The use of periodic cold-calling across the semester also served to diminish differences in participation levels between high and low participants. Additionally, students reported feeling nervous during cold-calling units, a finding in direct contrast to Dallimore et al. (2006) in which students reported no differences in classroom comfort.

There are many practical uses for cold-calling, though its implementation and use should be carefully considered. The type of population is perhaps the most pertinent to how cold-calling should be used. In a primary school setting, cold-calling can be used to encourage shy students to contribute more frequently. Cold-calling should be introduced at the beginning of the school year and used only with fact-based, simple questions at first. This will provide the students time to acclimate to the process and build confidence in speaking in front of the class. Having students discuss questions and responses in small groups may also help develop these skills. Teachers may wish to randomly call on students through the use of Popsicle sticks, student numbers, or other methods. Asking questions based on each student's individual strengths would minimize the possibility of social embarrassment.

In middle and high school settings, cold-calling should be implemented slightly differently. Though it would benefit students to introduce cold-calling earlier, teachers may choose to use more advanced questioning early on. While this may decrease students' comfort, it may lead to greater course preparation. Allowing an extended wait time after posing questions would provide students a chance to formulate answers.

Overall, it appears that a combination of voluntary and cold-calling could be most beneficial. Introducing cold-calling earlier to students may promote comfort within the

classroom and encourage students to prepare more for discussions. Students more inclined to participate would have the opportunity to thrive within a voluntary participation context, while students less inclined to participate would be encouraged to join classroom discussions within the cold-calling context. A greater balance in quantity of discussion across students would be the likely result.

List of References

Armstrong, M., & Boud, D. (1983). Assessing participation in discussion: An exploration of the issues. *Studies in Higher Education*, *8*(1), 33-44.

Bean, J. C., & Peterson, D. (1998). Grading classroom participation. *New Directions for Teaching and Learning, 74*, 33-40.

Boniecki, K. A., & Moore, S. (2003). Breaking the silence: Using a token economy to reinforce classroom participation. *Teaching of Psychology*, *30*(3), 224-227.

Brookfield, S. D., & Preskill, S. (2012). *Discussion as a way of teaching: Tools and techniques for democratic classrooms*. John Wiley & Sons.

Brown, C. T., & Pruis, J. J. (1958). Encouraging participation in classroom discussion. *The Speech Teacher, 7*(4), 344-346.

Burchfield, C. M., & Sappington, J. (1999). Participation in Classroom Discussion. *Teaching of Psychology*, *26*(4), 290-91.

Carstens, B. A., Wright, J. M., Coles, J. T., McCleary, L. N., & Williams, R. L. (2013). The role of self-monitoring in assessing individual students' quantity and quality of comments in large-class discussion. *Journal of Excellence in College Teaching, 24*(1), 123-146.

Chan, B., & McCroskey, J. C. (1987). The WTC scale as a predictor of classroom participation. *Communication Research Reports*, *4*(2), 47-50.

Constantinople, A., Cornelius, R., & Gray, J. (1988). The chilly climate: Fact or artifact? *The Journal of Higher Education*, *59*(5), 527-550.

Dallimore, E. J., Hertenstein, J. H., & Platt, M. B. (2004). Classroom participation and discussion effectiveness: Student-generated strategies. *Communication Education*, *53*(1), 103-115.

Dallimore, E. J., Hertenstein, J. H., & Platt, M. B. (2006). Nonvoluntary class participation in graduate discussion courses: Effects of grading and cold calling. *Journal of Management Education*, *30*(2), 354-377.

Dallimore, E. J., Hertenstein, J. H., & Platt, M. B. (2013). Impact of Cold-Calling on Student Voluntary Participation. *Journal of Management Education*, *37*(3), 305-341. (Arranged by year when author names are the same; pg. 182 of the APA manual)

Dancer, D., & Kamvounias, P. (2005). Student involvement in assessment: a project designed to assess class participation fairly and reliably. *Assessment & Evaluation in Higher Education*, *30*(4), 445-454.

Fassinger, P. A. (1995). Understanding classroom interaction: Students' and professors' contributions to students' silence. *The Journal of Higher Education*, *66*(1), 82-96.

Flanders, N. A. (1962). Using interaction analysis in the inservice training of teachers. *The Journal of Experimental Education*, *30*(4), 313-316.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research*, *74*(1), 59-109.

Fritschner, L. M. (2000). Inside the undergraduate college classroom: Faculty and students differ on the meaning of student participation. *Journal of Higher Education*, *71*(3), 342-362.

Galyon, C. E. (2012). *Analysis of the role of homework in predicting and improving exam performance*. (Doctoral dissertation). Retrieved from TRACE.

Galyon, C. E., Blondin, C. A., Yaw, J. S., Nalls, M. L., & Williams, R. L. (2012). The relationship of academic self-efficacy to class participation and exam performance. *Social Psychology of Education*, *15*(2), 233-249.

Garside, C. (1996). Look who's talking: A comparison of lecture and group discussion teaching strategies in developing critical thinking skills. *Communication Education, 45*(3), 212-227.

Gray, C. D., & Kinnear, P. R. (2012). *IBM SPSS Statistics 19 made simple*. New York, NY: Psychology Press.

Howard, J. R., & Henney, A. L. (1998). Student participation and instructor gender in the mixed-age college classroom. *Journal of Higher Education*, *69*(4), 384-405.

Jacobs, L. C., & Chase, C. I. (1992). *Developing and using tests effectively: A guide for faculty*. San Francisco, CA: Jossey-Bass.

Jones, R. C. (2008). The "why" of class participation: a question worth asking. *College Teaching*, *56*(1), 59-63.

Junn, E. (1994). “Pearls of wisdom": Enhancing student class participation with an innovative exercise. *Journal of Instructional Psychology, 21*(4), 385-387.

Karp, D. A., & Yoels, W. C. (1976). The college classroom: Some observations on the meanings of student participation. *Sociology & Social Research, 60*(40), 421-439.

Krohn, K. R., Foster, L. N., McCleary, D. F., Aspiranti, K. B., Nalls, M. L., Quillivan, C. C., Taylor, C. M., & Williams, R. L. (2011). Reliability of students’ self-recorded participation in class discussion. *Teaching of Psychology, 38,* 43-45.

Kuh, G. D., & Umbach, P. D. (2004). College and character: Insights from the national survey of student engagement. *New Directions for Institutional Research*, *2004*(122), 37-54.

Levin, J, & Fox, J. A. (2011). *Elementary Statistics in Social Research* (3rd ed.). Boston, MA: Pearson Education, Inc.

Ma, H. H. (2006). An alternative method for quantitative synthesis of single-subject researches: Percentage of data points exceeding the median. *Behavior Modification, 30*, 598-617.

McCorskey, J.C., & McVetta, R.W. (1978). Classroom seating arrangements: instructional communication theory versus student preferences. *Communication Education, 27*, 99-111.

Melvin, K. B. (1988). Rating class participation: The prof/peer method. *Teaching of Psychology*, *15*(3), 137-139.

Morrison, T. L., & Thomas, M. D. (1975). Self-esteem and classroom participation. *The Journal of Educational Research*, *68*(10), 374-377.

Murray, H., & Lang, M. (1997). Does classroom participation improve student learning. *Teaching and Learning in Higher Education*, *20*, 7-9.

Petress, K. (2006). An operational definition of class participation. *College Student Journal*, *40*(4), 821-823.

Rocca, K. A. (2010). Student participation in the college classroom: An extended multidisciplinary literature review. *Communication Education*, *59*(2), 185-213.

Scruggs, T. E., Mastropieri, M. A., Cook, S. B., & Escobar, C. (1986). Early interventions for children with conduct disorders: A quantitative synthesis of single-subject research. *Behavioral Disorders*, *11*, 260-271.

Sommer, R., & Sommer, B. A. (2007). Credit for comments, comments for credit. *Teaching of Psychology*, *34*(2), 104-106.

Spender, D. (1982). *Invisible women: The schooling scandal*. London: Writers and Readers.

Weaver, R. R., & Qi, J. (2005). Classroom organization and participation: College students' perceptions. *The Journal of Higher Education*, *76*(5), 570-601.

Williams, R. L. (1971). Relationship of class participation to personality, ability, and achievement variables. *The Journal of Social Psychology*, *83*(2), 193-198.

Zaremba, S. B., & Dunn, D. S. (2004). Assessing class participation through self-evaluation: Method and measure. *Teaching of Psychology*, *31*(3), 191-193.

Appendices

Appendix A: Course Grading Structure

Regular Credit

1. Attendance and Name Card Display (60 total points—up to 3 points per day for the first four days in each unit leading and up to 12 points per unit).

2. Class Participation (80 total points—up to 6 points per day, maximum 20 points per units B-E)

3. Practice Exams (25 total points—up to 5 points per exam)

4. Unit Exams (250 total points—up to 50 points on each of five unit exams)

5. Final Exam (100 total points)

6. Research Participation (15 possible points—up to 5 points for completing each of three research inventories)

Bonus Credit Opportunities

7. Cooperative learning bonus (10 possible points based on meeting the specified cooperative learning contingencies)

8. Mystery Day Bonus (5 points if you attended on the randomly selected Mystery Day and 5 additional points if you attended on all Mystery Days)

Grade Scale

You can earn a maximum of 530 points of regular credit in the course, exclusive of 20 bonus points. No credit options beyond those described in this syllabus will be offered. **Do not request any personal adjustment in the grading scale at any time during the course**. The grade scale is as follows:

A	=	90% and above	474 and above
B+	=	88-89%	464-473
B	=	80-87%	421-463
C+	=	78-79%	411-420
C	=	70-77%	368-410
D	=	60-69%	315-367
F	=	below 60%	314 and below

Appendix B: GTA Observation Form

Observer Participation Record

Year _____ Semester _____

Date_____

Observer _____

Tx Schedule _____ Section _____ Unit _____ Day_____ Class Time _____

Instructor_____

I = Ignored Comment
R = Restated Comment
P = Gave Positive Verbal Feedback
C = Corrected Comment
O = Asked for Other Thoughts

Name	Cold-Called	Voluntary

Appendix C: Instructions for Students

Treatment Instructions for Students

Voluntary Participation Unit

During this unit, participation will be strictly voluntary. If you would like to answer a question, ask a question, or express an opinion about an issue, you should raise your hand. I will first recognize the students who have commented the least number of times during the unit. It is important to remember that participation is part of your final grade. You must take the initiative to earn participation credit in this unit. All comments or questions made during this unit should be recorded and rated as "voluntary" comments.

Called-On Unit

During the called-on unit, I will ask a question first and then identify a student to answer the question. I will call on each of you in a predetermined random order. To be well prepared to answer instructor questions when called on, you must have answered all the instructor-notes, video, and article questions prior to class and then listen closely to every instructor question posed in class. If time permits, I will call on each of you three times in a class period. When called on, you should record your response and rating on the "called on" side of your record card. If you ask me to repeat a question, I will call on another student to answer the question. Asking me to repeat a question or saying you can't answer the question will count as one of your "called-on" opportunities and should be recorded as a 0-point comment. Your responses should be recorded and rated (0 to 2) as usual during this unit.

Although you are not to volunteer comments or opinions during this unit, you are free to ask questions. If you have a question about what the instructor or another student has just said or information in the course material related to what has just been said in class, you should raise

your hand to get the instructor's attention. When asking a question, you should record it on the "voluntary" comment side of your record card. Questions are to be rated in the same manner as comments. If you ask for information or an explanation I have already provided that day, you should circle a 0 for that question. If you ask for an explanation of some statement or concept in the discussion without first stating your understanding of that information, you would rate that question as a 1. On the other hand, if you ask a question by first stating your understanding of a particular point, you should rate that question as a 2.

If you have any questions about the procedures to be used in the called-on units, contact me by email or phone to get further clarification of the called-on procedures. I will spend minimal time in class reviewing these instructions. The purpose of comparing the two different ways of managing the discussion in this course is to determine the best way to give every student an opportunity to participate in the discussion and to maximize your understanding of the course material.

Appendix D: Tables

Table 1

Flow of Treatment across Sections and Units

	Section	Unit A	Unit B	Unit C	Unit D	Unit E
GTA 1						
	1	Baseline	Voluntary	Cold-Calling	Voluntary	Cold-Calling
	4	Baseline	Cold-Calling	Voluntary	Cold-Calling	Voluntary
GTA 2						
	2	Baseline	Cold-Calling	Voluntary	Cold-Calling	Voluntary
	3	Baseline	Voluntary	Cold-Calling	Voluntary	Cold-Calling
GTA 3						
	5	Baseline	Voluntary	Cold-Calling	Voluntary	Cold-Calling
	6	Baseline	Cold-Calling	Voluntary	Cold-Calling	Voluntary

Table 2.

Percentage of Inter-rater Agreement for Credit Ratings across All Sections and Units

	Units											
Section	B			C			D			E		
	S1	S2	Os	S1	S2	Os	S1	S2	Os	S1	S2	Os
1	69	69	100	92	92	100	83	82	100	93	93	100
2	88	88	100	70	70	91	85	84	99	81	81	96
3	86	86	100	82	85	96	81	79	97	88	80	91
4	89	90	96	82	78	96	94	93	99	90	90	100
5	80	78	98	91	92	96	86	88	98	89	89	100
6	90	89	98	75	73	94	88	86	96	80	83	96
Mean	84	83			82	96	86	85	98	87	86	97

Note. S1 = Student and Observer 1, S2 = Student and Observer 2, Os = Observer 1 and 2.

Table 3

Means for Capped Participation across Graduate Teaching Associates and Units

		Units						
	Section	A	B	C	D	E	Voluntary Average	Cold-Calling Average
GTA 1								
	1	13.50	16.46_V	14.35_C	15.92_V	14.62_C	16.21	14.48
	4	14.96	16.86_C	14.57_V	13.46_C	16.43_V	15.50	15.25
GTA 2								
	2	15.81	16.12_C	17.81_V	14.19_C	16.23_V	17.06	15.27
	3	14.54	14.93_V	12.54_C	15.93_V	12.71_C	15.50	12.82
GTA 3								
	5	13.56	14.88_V	13.60_C	16.12_V	9.56_C	15.60	14.26
	6	11.56	15.52_C	15.39_V	12.26_C	15.70_V	15.54	16.59
	Mean	13.99	15.80	14.71	14.65	14.21	15.90	14.78

Table 4.

Means for Raw Participation across Graduate Teaching Associates and Units

		Units						
	Section	A	B	C	D	E	Voluntary Average	Cold-Calling Average
GTA 1								
	1	15.38	18.73_{V}	15.31_{C}	18.15_{V}	14.58_{C}	18.44	14.94
	4	18.57	18.11_{C}	15.57_{V}	13.89_{C}	18.36_{V}	16.96	16.00
GTA 2								
	2	17.27	16.88_{C}	20.50_{V}	15.15_{C}	18.38_{V}	19.44	16.02
	3	15.29	15.57_{V}	12.93_{C}	17.21_{V}	13.54_{C}	16.39	13.23
GTA 3								
	5	14.88	16.04_{V}	13.88_{C}	17.00_{V}	9.56_{C}	16.52	11.72
	6	12.83	16.87_{C}	17.48_{V}	12.61_{C}	17.35_{V}	17.41	14.74
	Mean	15.79	17.04	15.88	15.72	15.31	17.52	14.46

Table 5.

Raw Participation Means for High- and Low-level Participants across Units B-D

	Participation Level	Mean	Std. Deviation	N
Voluntary Average	Low	12.05	6.97	40
	High	20.84	4.69	41
	Total	16.45	5.83	81
Cold-Calling Average	Low	12.63	3.67	40
	High	16.26	3.46	41
	Total	14.44	3.56	81

Table 6.

Interaction Effect of Participation Groups and Discussion Condition on Raw Participation Means

	High	Low	
Voluntary	20.89	12.08	16.59
Cold-calling	16.36	12.39	14.42
-	18.63	12.24	

Table 7.

Opportunities to Participate During Each Unit and Treatment Condition

		Units						
	Section	A	B	C	D	E	Voluntary Average	Cold-Calling Average
GTA 1								
	1	234	272_V	236_C	258_V	223_C	265	230
	4	293	324_C	240_V	230_C	274_V	257	277
GTA 2								
	2	240	242_C	277_V	204_C	247_V	262	223
	3	229	226_V	198_C	255_V	201_C	241	200
GTA 3								
	5	233	241_V	211_C	240_V	136_C	241	174
	6	225	260_C	240_V	169_C	229_V	235	215
	Mean	242	261	234	226	218	250	220

Table 8.

Participation Rate Mean Percentages across Participation Groups

	Participation Group	Mean %	Std. Deviation %	N
Voluntary Average	Low	2.44	1.36	41
	High	4.51	0.88	37
	Total	3.77	0.91	78
Cold-Calling Average	Low	3.29	0.92	41
	High	3.95	0.88	37
	Total	3.64	0.78	78

Table 9.

Means for Attendance across Graduate Teaching Associates, Units, and Treatment Conditions

		Units						
	Section	A	B	C	D	E	Voluntary Average	Cold-Calling Average
GTA 1								
	1	7.31	7.46_V	6.77_C	6.85_V	7.54_C	7.15	7.15
	4	7.71	7.21_C	7.00_V	6.57_C	7.29_V	7.14	6.89
GTA 2								
	2	7.46	7.31_C	7.69_V	7.00_C	7.23_V	7.46	7.15
	3	7.29	7.21_V	7.43_C	7.50_V	7.43_C	7.36	7.43
GTA 3								
	5	7.84	7.60_V	7.68_C	7.60_V	7.44_C	7.60	7.56
	6	7.74	7.83_C	7.57_V	7.65_C	7.74_V	7.65	7.74
	Mean	7.55	7.42	7.35	7.18	7.44	7.38	7.31

Table 10.

Means for Exam Scores across Graduate Teaching Associates and Units

		Units						
	Section	A	B	C	D	E	Voluntary Average	Cold-Calling Average
GTA 1								
	1	39.92	41.50_V	41.04_C	41.58_V	39.00_C	41.54	40.02
	4	40.93	40.89_C	39.96_V	41.54_C	39.00_V	39.48	41.21
GTA 2								
	2	44.58	37.35_C	47.62_V	42.65_C	40.04_V	43.83	40.00
	3	40.21	37.68_V	47.07_C	41.50_V	40.11_C	39.59	43.59
GTA 3								
	5	42.24	39.72_V	41.52_C	41.96_V	38.68_C	40.84	40.10
	6	40.61	37.48_C	41.22_V	41.00_C	38.74_V	39.98	39.24
	Mean	41.40	39.13	43.13	41.71	39.28	40.86	40.77

Table 11.

Means for Exam Z-Scores across Graduate Teaching Associates and Units

		Units						
	Section	A	B	C	D	E	Voluntary Average	Cold-Calling Average
GTA 1								
	1	0.09	0.72_V	0.29_C	0.20_V	0.02_C	0.46	0.15
	4	0.28	0.63_C	0.12_V	0.20_C	0.02_V	0.07	0.41
GTA 2								
	2	0.95	0.10_C	1.34_V	0.41_C	0.19_V	0.76	0.25
	3	0.15	0.15_V	1.25_C	0.19_V	0.20_C	0.17	0.73
GTA 3								
	5	0.52	0.45_V	0.36_C	0.28_V	0.04_C	0.37	0.16
	6	0.22	0.12_C	0.32_V	0.09_C	0.03_V	0.14	0.11
	Mean	0.36	0.37	0.62	0.23	0.06	0.33	0.31

Table 12.

Interaction Effect for Onset Condition and Discussion Condition across Exam Z-scores

	Unit A	Voluntary	Cold-calling	
Early Onset	0.49	0.32	0.27	0.36
Late Onset	0.25	0.33	0.36	0.31
	0.36	0.33	0.31	

Appendix E: Figures

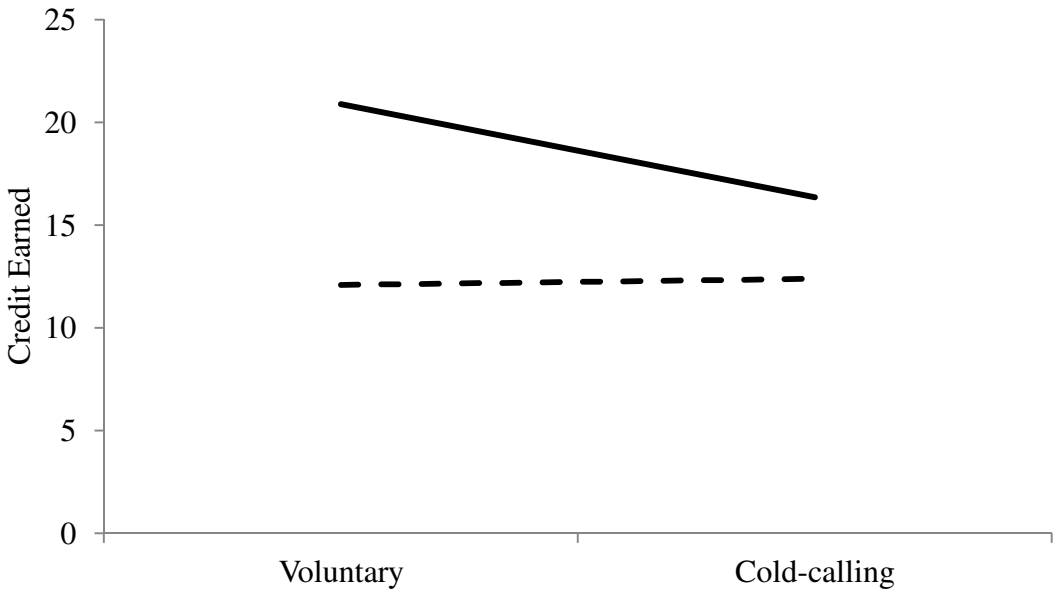

Figure 1. Interaction Effect of High- and Low-level Participants and Discussion Condition on Raw Participation Means

Note. The dotted line represents the low-level participants and the solid line represents the high-level participants.

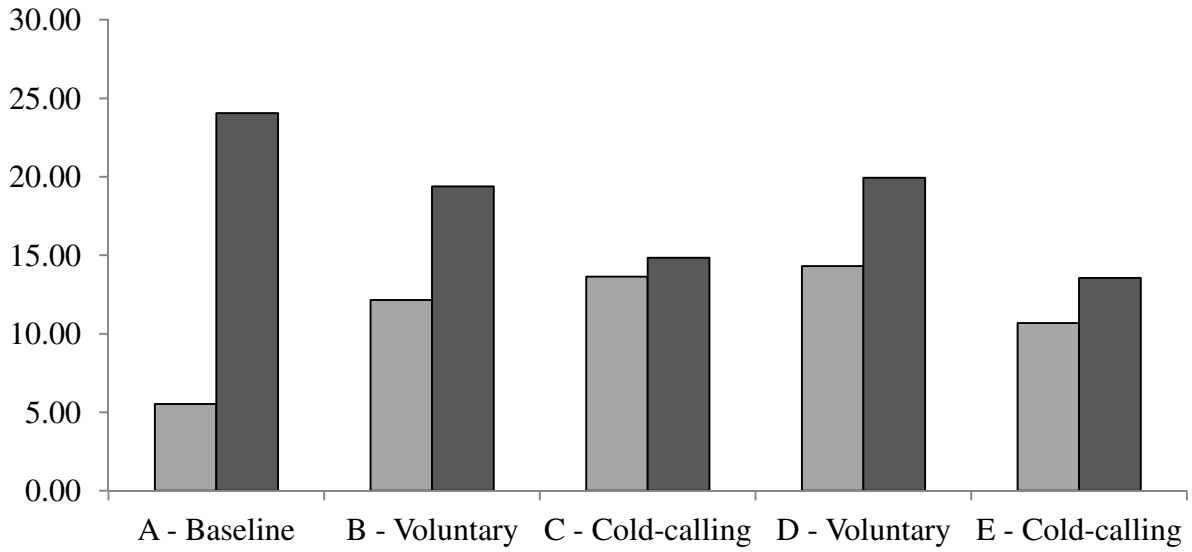

Figure 2. Participation Means for High and Low Participants in the Late Onset Condition Across Units

Note. The light gray column represents the low-level participants and the dark gray column represents the high-level participants.

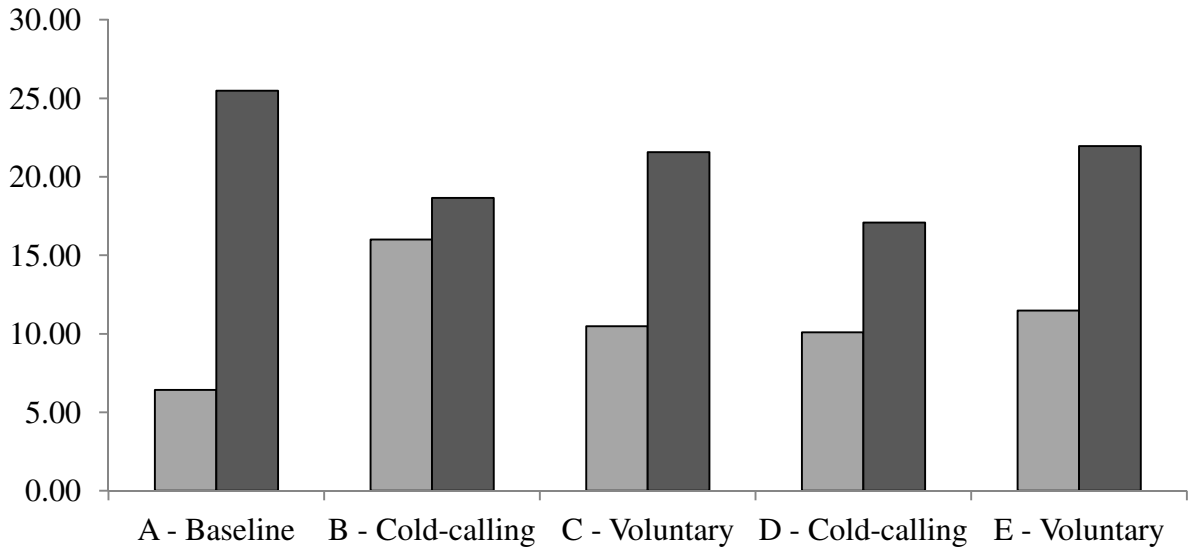

Figure 3. Participation Means for High and Low Participants in the Early Onset Condition across Units

Note. The light gray column represents the low-level participants and the dark gray column represents the high-level participants.

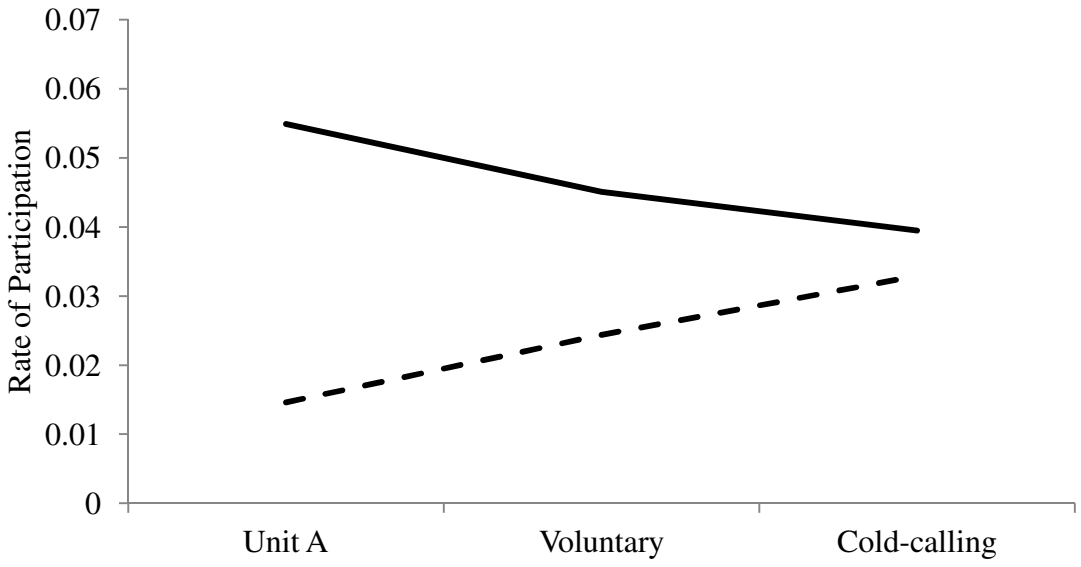

Figure 4. Interaction Effect of High- and Low-level Participants and Discussion Condition on Participation Rate

Note. The dotted line represents the low-level participants and the solid line represents the high-level participants.

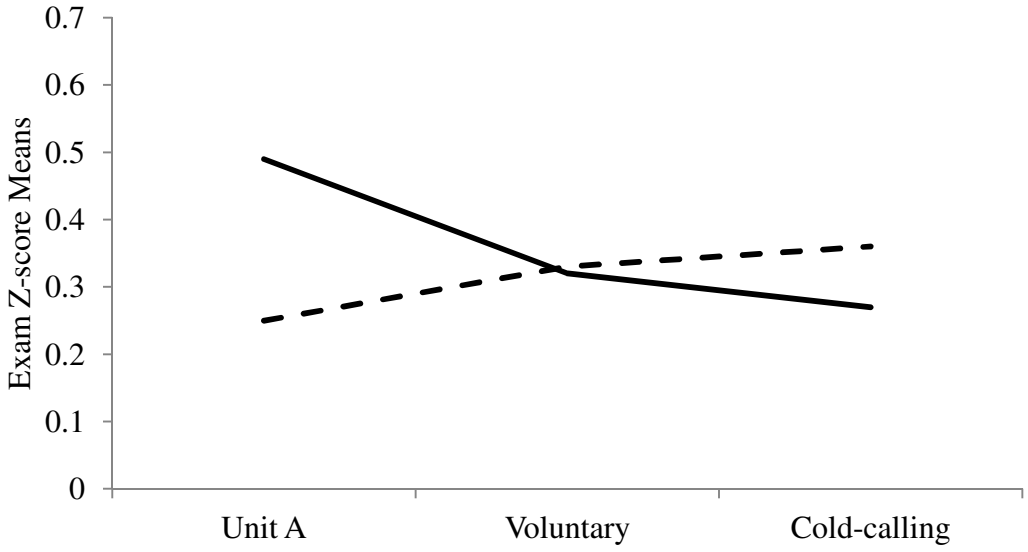

Figure 5. Interaction Effect of Early and Late Onset and Discussion Condition on Exam Z-score Means

Note. The dotted line represents the late onset condition and the solid line represents the early onset condition.

Appendix F: Participation Survey

The following questions pertain to your participation in the Ed Psych class discussions this semester. Please note that your responses will have no bearing on your final grade. Your answers are strictly intended for research purposes. If you wish to participate, you should put your answers on the scan form given you in class on the final exam review day. Returning your completed scan form when you take the final exam indicates you are willing to have your responses included in the research database. If you wish to receive the 5 bonus points for completing the survey, you will need to sign your name on the scan form and indicate what section you are in the space provided. Survey responses will automatically be scanned into the research database and no names will ever appear in the database. For the following statements, please indicate the degree to which you agree or disagree.

A = strongly agree B = agree C = neutral D = disagree E = strongly disagree

1) Overall, I earned more participation credit in the voluntary participation units than in the called-on units. (M = 2.15, SD = 1.29)

A) Strongly agree (46.2%)
B) Agree (16.6%)
C) Neutral (18.6%)
D) Disagree (13.1%)
E) Strongly Disagree (5.5%)

2) I regularly volunteered comments during the voluntary discussion units. (M = 1.75, SD = 1.05)

A) Strongly agree (55.9%)
B) Agree (24.8%)
C) Neutral (10.3%)
D) Disagree (6.2%)
E) Strongly Disagree (2.8%)

3) I found the instructor's responses to my comments generally helpful. (M = 1.52, SD = 0.65)

A) Strongly agree (55.2%)
B) Agree (37.9%)
C) Neutral (6.2%)
D) Disagree (0.7%)
E) Strongly Disagree (0.0%)

4) I understood how to get full participation credit during called-on units. (*M* = 1.82, *SD* = 1.06)

A) Strongly agree (51.0%)
B) Agree (29.0%)
C) Neutral (9.0%)
D) Disagree (9.0%)
E) Strongly Disagree (2.1%)

5) I felt well prepared to answer instructor question in the voluntary participation units. (*M* = 1.61, *SD* = 0.71)

A) Strongly agree (50.0%)
B) Agree (41.0%)
C) Neutral (6.9%)
D) Disagree (2.1%)
E) Strongly Disagree (0.0%)

6) I felt well prepared to answer instructor questions during called-on units. (*M* = 2.52, *SD* = 1.05)

A) Strongly agree (14.6%)
B) Agree (42.4%)
C) Neutral (23.6%)
D) Disagree (15.3%)
E) Strongly Disagree (4.2%)

7) I felt that most students were well prepared for class discussion during called-on units. (*M* = 2.61, *SD* = 1.00)

A) Strongly agree (10.3%)
B) Agree (40.7%)
C) Neutral (31.7%)
D) Disagree (12.4%)
E) Strongly Disagree (4.8%)

8) I generally would have preferred a lecture format to either of the discussion formats. (*M* = 3.22, *SD* = 1.28)

A) Strongly agree (15.2%)
B) Agree (12.4%)
C) Neutral (22.8%)
D) Disagree (34.5%)
E) Strongly Disagree (15.2%)

9) I prepared more for called-on participation units than voluntary units. (M = 2.52, SD = 1.20)

A) Strongly agree (23.4%)
B) Agree (29.0%)
C) Neutral (26.2%)
D) Disagree (14.5%)
E) Strongly Disagree (6.9%)

10) I generally listened closely to other students' comments in class discussion. (M = 1.98, SD = 0.85)

A) Strongly agree (27.6%)
B) Agree (55.2%)
C) Neutral (9.7%)
D) Disagree (6.9%)
E) Strongly Disagree (0.7%)

11) My instructor helped to make the called-on units comfortable. (M = 1.72, SD = 0.86)

A) Strongly agree (50.3%)
B) Agree (31.7%)
C) Neutral (14.5%)
D) Disagree (2.8%)
E) Strongly Disagree (0.7%)

12) I felt nervous during called-on units. (M = 2.01, SD = 1.12)

A) Strongly agree (41.4%)
B) Agree (32.4%)
C) Neutral (13.8%)
D) Disagree (8.3%)
E) Strongly Disagree (4.1%)

13) My instructor followed a strictly random procedure in calling on students during called-on units. (M = 1.43, SD = 0.73)

A) Strongly agree (67.6%)
B) Agree (24.1%)
C) Neutral (6.2%)
D) Disagree (1.4%)
E) Strongly Disagree (0.7%)

14) I understood how to get full participation credit during voluntary units. (*M* = 1.36, *SD* = 0.63)

A) Strongly agree (71.0%)
B) Agree (23.4%)
C) Neutral (4.1%)
D) Disagree (1.4%)
E) Strongly Disagree (0.0%)

15) I generally enjoyed sharing my perspectives on course issues. (*M* = 2.29, *SD* = 0.96)

A) Strongly agree (22.8%)
B) Agree (37.9%)
C) Neutral (27.6%)
D) Disagree (11.0%)
E) Strongly Disagree (0.7%)

16) I felt comfortable volunteering questions during called on units. (*M* = 2.70, *SD* = 1.10)

A) Strongly agree (14.5%)
B) Agree (31.0%)
C) Neutral (30.3%)
D) Disagree (18.6%)
E) Strongly Disagree (5.5%)

17) Overall, the instructor managed the called-on procedure in the way it had been explained to the class. (*M* = 1.55, *SD* = 0.76)

A) Strongly agree (57.2%)
B) Agree (33.8%)
C) Neutral (6.2%)
D) Disagree (2.1%)
E) Strongly Disagree (0.7%)

18) The students were generally respectful of others' comments. (*M* = 1.46, *SD* = 0.68)

A) Strongly agree (62.1%)
B) Agree (32.4%)
C) Neutral (4.1%)
D) Disagree (0.7%)
E) Strongly Disagree (0.7%)

19) I found it difficult to keep track of the class discussions in the voluntary units. (M = 3.95, SD = 0.89)

A) Strongly agree (2.8%)
B) Agree (4.1%)
C) Neutral (13.1%)
D) Disagree (55.2%)
E) Strongly Disagree (24.8%)

20) I followed the discussion more closely during called- on units. (M = 2.75, SD = 1.15)

A) Strongly agree (14.5%)
B) Agree (30.3%)
C) Neutral (28.3%)
D) Disagree (19.3%)
E) Strongly Disagree (7.6%)

21) My instructor appeared to enjoy the called-on units. (M = 2.28, SD = 1.07)

A) Strongly agree (24.8%)
B) Agree (42.1%)
C) Neutral (16.6%)
D) Disagree (13.8%)
E) Strongly Disagree (2.8%)

22) The called-on units increased the probability that everyone would participate in class discussion. (M = 2.81, SD = 1.34)

A) Strongly agree (19.3%)
B) Agree (28.3%)
C) Neutral (19.3%)
D) Disagree (18.6%)
E) Strongly Disagree (14.5%)

23) I feel that class discussion should be managed strictly on a voluntary basis. (M = 2.06, SD = 1.16)

A) Strongly agree (47.6%)
B) Agree (15.9%)
C) Neutral (20.0%)
D) Disagree (16.6%)
E) Strongly Disagree (0.0%)

24) The called-on procedure felt too regimented to me. (M = 1.90, SD = 1.13)

A) Strongly agree (52.4%)
B) Agree (20.7%)
C) Neutral (13.1%)
D) Disagree (12.4%)
E) Strongly Disagree (1.4%)

25) Overall, I did better on exams in the called-on units. (M = 3.39, SD = 1.08)

A) Strongly agree (4.1%)
B) Agree (14.5%)
C) Neutral (39.3%)
D) Disagree (22.8%)
E) Strongly Disagree (19.3%)

Please review your scan form to be sure you responded to each item.

Vita

Brittany Ann Carstens was born at Yokota Air Base in Tokyo, Japan. She moved to the Mississippi Gulf Coast as a young child, where she spent her childhood and adolescence. She obtained a B.A. in Psychology from the University of Mississippi in Oxford, Mississippi in 2010. She then enrolled in the University of Tennessee's School Psychology Ph.D. program later that year. In December of 2012, she received her M.S. in Applied Educational Psychology. Brittany will receive her Ph.D. in August 2015 upon completion of a year-long pre-doctoral internship with the Tennessee Internship Consortium in Knoxville, TN.