University of Tennessee, Knoxville
Trace: Tennessee Research and Creative Exchange

Doctoral Dissertations | Graduate School

8-2011

Inter-rater Reliability and Related Variables of a Newly Developed Measure of Quality of Student Participation

Lisa Nicole Edge
lisaedge3@gmail.com

Recommended Citation
Edge, Lisa Nicole, "Inter-rater Reliability and Related Variables of a Newly Developed Measure of Quality of Student Participation. " PhD diss., University of Tennessee, 2011.
http://trace.tennessee.edu/utk_graddiss/1074

This Dissertation is brought to you for free and open access by the Graduate School at Trace: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of Trace: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Lisa Nicole Edge entitled "Inter-rater Reliability and Related Variables of a Newly Developed Measure of Quality of Student Participation." 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:

Chris H. Skinner, Ralph S. McCallum, David F. Cihak

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

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

Inter-rater Reliability and Related Variables of a Newly Developed Measure of Quality of Student Participation

A Dissertation
Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Lisa Nicole Edge
August 2011

Copyright © 2011 by Lisa Nicole Edge
All rights reserved.

Acknowledgements

First and foremost, I am most grateful for my educational experiences in the psychology program at Castleton State College, especially Dr. Bergen, whose unorthodox homework assignments and class discussions instilled in me the appropriate attitudes of a future scientist and critical thinker, as well as applied statistical skills. My sister, Lauren Edge, who has overcome many difficult challenges in her life with grace, is my hero. Her friendship and strength is an inspiration to me and motivated me to remain strong and confident in my weakest moments during my educational journey and the completion of my dissertation.

I would also like to thank the members of my committee, Drs. David Cihak, Chris Skinner, and Steve McCallum, whose suggestions and constructive criticisms regarding methodological design, research implications, and written expression contributed significantly to the completion of this project and my professional development. Also, I am very grateful for Dr. Robert Williams, who is a wonderful mentor and educator. His tremendous work ethic, value of critical thinking, professionalism, and sense of humor will always be an inspiration to me and has shaped me into the professional that I am today. I am especially grateful for Drs. Skinner, McCallum, and Williams for going above and beyond the call of duty as educators during a time of need the very year that my dissertation data were collected.

I would also like to thank Cora Taylor, who was gracious enough to allow me to observe all discussions that took place in both class sections that she taught. Finally, I would like to thank Daniel McCleary, who not only assisted in the development of the coding scheme used in this study and data collection, but is also an outstanding friend. His moral character is an

inspiration to me and to all who meet him. Without his love, encouragement, support, and ability to laugh, my graduate school experience would have been far less enjoyable.

Abstract

Some researchers suggest that quality is preserved when students are rewarded for quantity of class participation (Boniecki & Moore, 2003; Bruss, 2009; Mainkar, 2007; Zaremba & Dunn, 2004); however, few studies have targeted the systematic assessment of participation quality. The primary purpose of the study was to develop a reliable system for rating quality of student participation, investigate whether quality of participation is preserved when students are given credit for the amount of class participation, and examine the relationship between participation quality and important course variables.

The researcher in the current study developed a rating system to evaluate the quality of student participation in 2 small sections of an undergraduate class at a large Southeastern university. The primary observer rated the quality of each student comment and recorded the number of comments each student contributed each discussion day. In order to assess the reliability of the coding scheme, a secondary observer rated the participation quality and quantity on the third day in each unit. As outlined in the syllabus, instructors awarded credit for the amount of participation on randomly drawn days at the conclusion of select units.

The average inter-observer agreement was 90% for the number of productive comments contributed by each student and 49% for the number of non-productive comments contributed by each student. The percent of productive comments that each student contributed each day was the primary dependent variable. Visual inspection and proportion analyses of the percent of productive comments revealed that quality of participation was generally preserved during units in which credit was awarded for the frequency of participation. On average, students who participated frequently were significantly more productive than those who participated

infrequently. Additionally, a student's frequency of course participation and critical thinking at the onset of the course significantly predicted classification into high- and low-quality responders. Future research suggestions include the following: expanding the definition of the quality of student participation into 3 overall qualitative categories rather than 2, increasing the criteria for participation credit in small course sections, and providing credit for the quality of participation.

Table of Contents

Chapter I 1
Introduction and Literature Review 1
Optimal Conditions for Participation 1
Preserving Quality While Targeting Quantity of Participation 2
Assessing the Quality of Student Participation 3
Participation History and Beliefs 6
Critical Thinking and Quality of Participation 7
Framework for the Current Study 8
Research Objectives 9
Chapter II 10
Method 10
Participants 10
Course Structure 10
Participation Contingencies 15
Student Participation Survey 16
Assessment of Critical Thinking 17
Unit Exams 17
Participation Contingencies and Research Design 17
Chapter III 18
Results 18
Descriptive and Correlation Analyses 18
Reliability 19
The Effect of Credit Contingencies on Participation 25
Intra-Student Comparisons of Frequency of Comments 30
Predicting Consistent Inclusion in a Qualitative Category 32
Differences among Qualitative Levels within Each Unit 38
Chapter IV 42
Discussion 42
Reliability of the Participation Quality Measure 42
Relationships between Course Variables and Participation Quality 49
Secondary Research Findings 51
Limitations and Directions for Future Research 58
List of References 63
Appendices 70
Appendix A 71
Appendix B 92
Appendix C 104
Appendix D 105
Appendix E 109
Appendix F 110
Appendix G 113
Appendix H 116
Vita 127

List of Tables

Table 1 Treatment Conditions Assigned Across the Two Sections 71
Table 2 Mean and Standard Deviations of Daily Percent of Productive Comments Based on Student Gender and Academic Classification 72
Table 3 Mean and Standard Deviations of Number of Comments Students Contributed Based on Student Gender and Acadmeic Performance 73
Table 4 Percent of Agreement between Primary and Secondary Raters for Productive Categories 74
Table 5 Point-Biserial Correlations Representing Inter-Rater Agreement for Productive and Nonproductive Comments 76
Table 6 Percent of Agreement between Primary and Secondary Raters for Non-Productive Comments 77
Table 7 Average Inter-Observer Percent Agreement between Student and Primary Observer Records of Number of Comments on Baseline (b), Credit (c), and Non-Credit (nc) Units 79
Table 8 Average Inter-Observer Percent Agreement (IOA) between Student and Observer Records of Number of Comments on Inter-rater Check in Baseline (b), Credit (c), and Non-Credit (nc) Units 80
Table 9 Differences in Proportions of Students at Different Qualitative Levls between Treatment Conditions Overall and between Pairs of Adjacent Treatment Units in Each Section 81
Table 10 Differences in Proportions of Productive Comments Made Each Discussion Session between Treatment Conditions Overall and between Pairs of Adjacent Treatment Units in Each Section 82
Table 11 Differences in Proportions of Students at Different Quantitative Levels between Treatment Conditions Overall and between Pairs of Adjacent Treatment Units in Each Section 83
Table 12 Number of Low-Responding Students in Unit 1 Who Fell into Low, Medium, and High Categories in Subsequent Units 84
Table 13 Number of Medium-Responding Students in Unit 1 Who Fell into Low, Medium, and High Categories in Subsequent Units 85
Table 14 Number of High-Responding Students in Unit 1 Who Fell into Low, Medium, and High Categories in Subsequent Units 86
Table 15 Mean Scores and Standard Deviations on the Participation Survey for Low - and High-Quality Participating Students 87
Table 16 Mean Scores and Standard Deviations on Unit Exams for the Unit's Low- and High-Quality Participating Students 88
Table 17 Mean Scores and Standard Deviations of Average Participation Frequency per day for each Unit's Low- and High-Quality Participating Students 89
Table 18 Discriminant-Function Analysis Results of Course Variables and Qualitative Groups 90
Table 19 The Average Proportion of Students at Each Participation Level within each Qualitative Level Overall and across Units 91

List of Figures

Figure 1. Record card for recording participation and daily credit activities..... 93
Figure 2. Percent of D-Level responders (67% or less productive) each day and A-Level responders (99 to 100% productive) each day..... 94
Figure 3. Percent of C-Level responders (68% - 88% productive comments) each day..... 95
Figure 4. Percent of B-Level responders (89% - 98% productive comments) each day..... 96
Figure 5. Percent of productive comments each discussion day..... 97
Figure 6. Percent of non-responders (0 comments) each day. 98
Figure 7. Percent of occasional responders (1-3 comments) each day..... 99
Figure 8. Percent of frequent responders (4-6 comments) each day. 100
Figure 9. Percent of dominating responders (7+ comments) each day. 101
Figure 10. Percent of low-responding students in Unit 1 who participated in subsequent units. 102
Figure 11. Overall mean percent of productive comments for high and low responders across units and sections. 103

Chapter I

Introduction and Literature Review

Many instructors acknowledge that active student participation is an integral factor in success of student learning. Approximately 93% of course syllabi included class participation as a contributor to a student's final grade at Seattle University (Bean & Peterson, 1998); however, instructor assessment of student participation is often subjective (Zaremba & Dunn, 2004). Although there are many recommendations to include multiple measurements of participation (i.e., quantity and quality) to increase objectivity and diminish subjectivity, few studies have targeted the systematic assessment of participation quality (Mainkar, 2007; Petress, 2006)

Optimal Conditions for Participation

Many factors have been related to certain aspects of student participation. In order to optimize student participation, Howard, James, and Taylor (2002) recommended that instructors convey to students that participation is valued and desired, provide students with a positive setting in which sharing of ideas can occur in a nonthreatening environment, and express to students that their experience and knowledge is beneficial to the class as a whole. Hodge and Nelson (1991) attempted to balance participation by reinforcing initially low responders with check marks on the chalkboard under the students' corresponding initials and ignoring dominant responders. The differential reinforcement procedure appeared to be useful in increasing participation from low responders and decreasing participation from dominant responders. It is important that students perceive that class discussion is their responsibility in order to increase the degree of student participation (Howard & Henney, 1998). Furthermore, Auster and MacRone (1994) suggested that instructors have a positive influence on student participation by calling on students by name, giving positive feedback for comments, and providing time for

students to answer. In addition, when students are given an opportunity to discuss course concepts in small groups, their ability to accurately answer higher order questions is greater than when students are only exposed to traditional lecture and cold calling (i.e., calling on students to answer specific questions without forewarning) methods (Garside, 1996). Thus, increased opportunity for students to discuss course content may optimize their critical thinking skills in that subject matter (Garside).

Auster and MacRone (1994) claim that students participate more in classes in which instructors ask predominantly higher order questions. Bloom's (1956) taxonomy of learning provides a framework from which to consider the complexity of instructor questions. The taxonomy has six categories: knowledge, comprehension, application, analysis, synthesis, and evaluation. However, Barnes (1983) found that only 18% of college instructors' questions were above Bloom's lowest level of learning. In conclusion, students may benefit most from instructors asking students to apply, compare, and contrast course concepts (Auster & MacRone; Barnes).

Preserving Quality While Targeting Quantity of Participation

One concern is that assigning credit for the amount of participation may cause students to participate purely for the sake of earning credit, potentially diminishing the quality of discussion (Gilson, 1994). For example, in a study conducted by Sommer and Sommer (2007), students reported that credit for participation would result in an increase of comments with no substance; yet when students actually received credit for participation, the majority (i.e., 71%) indicated that the quality of discussion actually improved. Additionally, Boniecki and Moore (2003) reported anecdotal evidence suggesting that the quality of participation remained consistent across phases when students were rewarded with tokens for the frequency of their class participation.

Although some evidence suggested that quality of participation is preserved even when students receive external rewards for the quantity of classroom contributions, researchers have yet to assess quality of participation in a systematic and comprehensive way. Researchers and students continue to fear the effects of giving students credit for frequency of participation on the quality of participation. Some researchers have designed mechanisms to buffer these potential adverse effects on quality. For example, Junn (1994) offered credit for the number of comments students contributed over a semester and offered bonus credit for comments that demonstrated a higher level of thinking, with the latter intended to minimize mundane comments. Unfortunately, Junn did not systematically assess the quality of student comments; therefore, the effect of the bonus points on discussion quality could not be ascertained. Rather, students and teachers anecdotally reported preservation of discussion quality despite increases in student participation.

Assessing the Quality of Student Participation

Many researchers have designed reliable methods to assess the frequency of student comments, yet few have examined objective methods of assessing the quality of student participation. Given that students are more comfortable when criteria for participation are known prior to the onset of the course, it is important to clearly define quality of participation at the outset of a course (Bruss, 2009; Stitt, Simonds, & Hunt, 2003). Quality of student participation has typically been assessed through the following measures: self-ratings, peer-ratings, and instructor ratings. Zaremba and Dunn (2004) investigated daily student self-rating of quality of participation using a single four-point Likert item. However, the reliability of student self-rating of quality of their own participation was not investigated. Zaremba and Dunn

reported that students' perceived increased responsibility for discussion quality when they completed the daily self-rating.

Peer evaluation of quality of student participation has been investigated by several researchers (Gilson, 1994; Gopinath, 1999; Mainkar, 2007; Melvin, 1988). Both peer-rated and self-rated participation have resulted in higher ratings of participation quality than have instructor ratings (Gopinath). Peer ratings of participation also provided a more reliable measure than self-ratings but were a poor predictor of instructor evaluations (Burchfield & Sappington, 1999, Ryan, Marshall, Porter, & Jia, 2007). Furthermore, students' attitude toward rating their own peers was typically negative (Gilson; Gopinath; Love, 1981).

Some researchers have used criterion-based rubrics to evaluate students' verbal participation in the classroom (Bruss, 2009; Mainkar, 2007; Ryan et al., 2007). Mainkar evaluated comments based on the following dimensions: no-substance (i.e., participation without contribution), straightforward, and insightful. Using a rubric that defined all of the dimensions, peer ratings achieved adequate inter-rater agreement in both semesters of the study (i.e., $r = .89$ and .90). No-substance comments included those that did not add to the understanding of the topic, repeated what another student said earlier, and were irrelevant to the topic being discussed. Straightforward comments included those that added to the understanding of the topic, adequately answered a question, and provided theory and linkages to what others said. Insightful comments included those that made the following contributions: improved the understanding of the topic, demonstrated depth and complexity of thought; or made a connection to theory or prior course concepts and took the discussion in a new direction. Petress (2006) proposed four dimensions of classroom participation that detract from productive class discussions: (1) long-winded/irrelevant contributions, (2) repetitive responses, (3) monopolization of participation, and

(4) responses that discourage others from contributing. Irrelevancy and repetition were recognized as two dimensions that diminish class discussion quality.

Bruss (2009) used an evaluation guide that classified students as A Contributors, B Contributors, C Contributors, D-F Contributors, and Non-responders. The distinction between these levels related to the frequency of substantive comments, ability to substantiate disagreements, the degree to which the student seemed prepared for class discussion, and degree to which the quality of class discussion would decrease if the student was not present. Bruss gave instructors the freedom to develop their own methods of evaluating participation. Some professors chose to evaluate students by jotting down notes describing each student's comments, while others developed rating systems (i.e., number of comments supported by course materials, number of coherent comments). Although the faculty and students reported that they found the method used to assess and develop student participation helpful in improving class discussion, Bruss did not systematically assess the quality of student comments.

Students in the Krohn (2010) study self-rated their comments as either relevant or non-relevant. Results indicated that students made few non-relevant comments according to their own ratings, as well as those of secondary observers. Furthermore, crediting the frequency of student participation did not affect the frequency of irrelevant student comments. Similarly, students in the Aspiranti (2010) study self-rated their comments as either timely or repetitious. Aspiranti identified few incidents of repetitious comments; again, credit for the frequency of student comments did not increase or decrease repetitious comments. Although inter-observer agreement was considered adequate, student-observer agreement regarding non-relevant and repetitious comments was low (Aspiranti; Krohn). In order to make judgments regarding the relevancy of student participation, raters must be very familiar with course concepts. Observers

who have taught the course, as was the case in both the Aspiranti and Krohn studies, should be better able to judge relevancy than students who are taking the course. Additionally, the infrequent incidents of irrelevant and repetitious comments may suggest the need for a more precise distinction between relevant and non-relevant comments.

Researchers have provided several tips for assessing the quality of participation. Bean and Peterson (1998) suggested that instructors collaborate with their students to create a rubric for scoring class participation that is specific to the context of the class. Bruss (2009) suggested that rating forms used to assess student participation in the classroom be closely aligned with the criterion-based rubrics used to assign students an overall grade. Training raters and students (e.g., rating modeled student speeches, discussing distinctions between qualitative criteria) rather than merely distributing the rating criteria increases consistency between observer and student ratings (Stitt et al., 2003). Thus, training raters results in higher reliability than simply distributing a description of the rating system.

Participation History and Beliefs

Krohn (2010) administered a 50-item survey designed to measure student attitudes towards participation. She performed several principal component factor analyses with varimax rotation that ultimately resulted in three factors: 1) Expectation of Discussion in College Courses, 2) Personal Benefits of Participation, and 3) Personal History and Confidence Regarding Participation. Krohn found significant differences among low, medium, and high responders for the first two factors and between high and low responders for the third factor.

Aspiranti (2010) administered the same 50-item survey Krohn (2010) used. In contrast to Krohn, Aspiranti's factor analyses yielded seven factors: 1) Personal History and Preference Regarding Class Participation, 2) Impact of Discussion on Course Value and Grades, 3)

Cognitive and Affective Investment in Class Discussion, 4) Relevance of Discussion, 5) Possible Impediments to Discussion, 6) Responsibility for Discussion, and 7) High-Quality Contribution to Discussion. Aspiranti indicated significant differences in Personal History and Preference Regarding Class Participation among low-, medium-, and high-participation groups. Aspiranti also found significant differences in Impact of Discussion on Course Value and Grades as well as Possible Impediments to Discussion between low- and high-participation groups. Aspiranti reported relationships similar to those identified by Krohn between survey factors and participation levels. Using a logistic regression analysis, Aspiranti and Krohn examined whether the survey factors significantly predicted participation levels. Aspiranti (2010) and Krohn (2010) found that History and Preference was the strongest predictor of classification into high- and low-participation groups in the overall course and across units. Although both studies produced significant differences in participation levels for common factors, Krohn extracted a more parsimonious factor structure than Aspiranti. Consequently, I used the factors identified by Krohn in the current study.

Critical Thinking and Quality of Participation

Higher order thinking can be viewed as a cognitive parallel to high-quality student participation. Smith (1977) demonstrated significant positive relationships among critical thinking and participation, encouragement, and peer-to-peer interaction. McCleary, Foster, and Williams (2010) found that high-frequency responders had higher critical thinking scores than low-frequency responders. Several researchers also have established that critical thinking is significantly and positively related to exam scores (Wallace & Williams, 2003; Williams, Oliver, Allin, Winn, & Booher, 2003a, 2003b; Williams, Oliver, & Stockdale, 2004; Williams & Worth, 2002).

Williams et al. (2003b) noted that students who earned high course grades, yet initially scored low on critical thinking, made greater critical thinking gains in the course than those earning low course grades and also initially scoring low on critical thinking. Williams and Stockdale (2003) sought to investigate differences in cognitive and study-habits variables between high-performing low critical thinkers (HPLCT) and low-performing low critical thinkers (LPLCT). They determined that HPLCT achieved higher practice and exam scores, as well as higher GPA, than LPLCT. Additionally, HPLCT were better note-takers and made greater gains in critical thinking during the course than either of the two comparison groups.

Framework for the Current Study

This study is a follow-up to the Krohn (2010), Krohn et al. (2010) and Aspiranti (2010) studies. Some procedures used in these studies were also used in the current study. The current study used the specially designed student record card that has been demonstrated to yield high submission rates, as well as accuracy and ease in assigning students daily credit for a variety of dimensions including daily participation (Foster et al., 2009).

The current study addresses some limitations of the both the Aspiranti (2010) and Krohn (2010) studies. Both studies attempted to assess a dimension of quality of participation (i.e., relevancy and timeliness); however, these dimensions proved too narrow to yield adequate reliabilities or trends corresponding to credit contingencies. Inadequate inter-rater reliability in the Aspiranti and Krohn studies between students and raters may have been due to very infrequent percentages of repetitious comments and/or students may not have understood the distinctions between timely, relevant, and repetitious comments well enough to make accurate classifications. A student comment in the Aspiranti and Krohn studies could be considered timely or relevant, even if it was actually inaccurate (presuming the instructor did not address the

inaccuracy of the comment). In order to systematically assess linkages between quality and other important measures, I broadened the definition of quality of participation by categorizing all non-productive comments and creating multiple levels of productive comments (e.g., answering comprehension versus providing a testimony). Furthermore, students did not make judgments regarding the quality of their participation. Rather, trained observers rated the quality of student participation.

Research Objectives

The primary research objective was to develop a reliable and valid measure of quality of participation. At this early stage of developing the qualitative measure, students were not asked to rate the quality of their participation. Instead, quality of participation was rated only by observers who had taught the course. In addition to the assessment of the inter-rater reliability of the qualitative measure, the researcher assessed the linkage between the qualitative measure and a variety of other course measures: quantity of participation, attitudes toward participation, critical thinking, and exam performance. Inasmuch as the design of the study did not permit cause-effect inferences regarding relations between quality of participation and other identified variables could not be determined, a discriminant-function analysis was used in assessing the relationships between quality of participation and the comparative measures. Investigating relationships between quantity of participation, attitudes toward participation, critical thinking, and exam performance can inform instructors and future research in increasing the quality of student participation.

Chapter II

Method

Participants

Students enrolled in two sections of an undergraduate psychology course (n = 46) at a large, Southeastern university during the fall semester of 2009 participated in this study. Students typically enroll in the course to fulfill a requirement for the University's Teacher Preparation Program. The same Graduate Teaching Assistant (GTA) taught both sections of the course, with approximately 25 students in each section. The students represented varying levels of education: 4.3% freshman, 39.1% sophomores, 32.6% juniors, 17.4% seniors, and 6.5% graduate students. The majority of the students were female (n = 31). Participants were awarded 5 points, approximately 1% of course credit, for their participation in the study and were given the option to obtain this credit in an alternative manner (i.e., five-page research paper). No students chose to earn the research credit through the alternative assignment.

Course Structure

The course consisted of five units (i.e., physical, cognitive, social, psychological, and values development). Three of the five units (i.e., Units 1, 3, and 5) were comprised of 5 days, while the remaining units (i.e., Units 2 and 4) required one extra day due to complexity of information in those units. A video was presented on the first day of each unit. Following the video, the remainder of the first class and the next 3 days of each unit involved discussion of the information presented in the instructor notes and PowerPoint slides available to students. The instructor notes and PowerPoint slides were a compilation of concept descriptors, facts, definitions, visual representations of concepts, and information gleaned from textbooks, peer-reviewed articles, and other resources. On each day students discussed the instructor notes, they

were required to complete a predetermined set of questions over the reading assigned for the day. Thus, students were expected to be prepared to discuss concepts in class. On the fourth or fifth day of each unit, the class took an article quiz, received feedback on their practice exam, and discussed the articles previously assigned for students to read. On the final day of each unit, the students completed a multiple-choice exam over the unit. The current study focused on the first 4 days of each unit, in which discussion centered on concepts embedded in the video, instructor notes, slides, and selected articles.

Student participation was tracked on the first 4 days of each unit in two consecutive evening sections of the course. Each section met for 75 minutes twice a week. The same Graduate Teaching Assistant (GTA) was the instructor for both courses, thus minimizing variability in student participation due to instructor differences. The supervising instructor of all sections of the course trained the GTA in methods of leading class discussion. The GTA was trained to ask more comprehension than factual questions. Additionally, the GTA provided feedback to students following their comments that summarized the information contained in the student comment, connected the comment to important course concepts or other student comments, and/or affirmed the importance of the student's comment. Comprehension questions were those for which the answers were not directly stated in the students' readings. Rather, students were required to compare/contrasts concepts or explain claims found in the readings. Alternatively, factual questions had answers that could be found directly in the students' readings. Both sections were identical in course proceedings, syllabi, content, assignments, materials, and schedule.

Participation Recording Procedures

The primary rater of the quality of student comments had taught the course for four semesters. Using the operational definitions outlined in Appendix C, the primary observer classified each student comment under the appropriate category on the first 4 days in each unit. Thus, the primary rater provided data to verify the reliability of student self-recording for the quantity dimension, as well as a measure of comment quality for all 4 days in each unit. Each of the 15 categories was represented as either a productive or a non-productive comment (see Appendix D). There were eight productive categories and seven nonproductive categories in which a comment could be classified. Productive and non-productive categories were analogous to the distinction between participation contribution and participation without contribution used in the Mainkar (2007) study. The following are comments included in the productive category: asking questions or offering perspectives that extend ideas in the notes, answering a factual question for which the answer is not found in course materials, accurately addressing a factual question in one's own words when the answer is found in course materials, questioning perspectives included in the course materials, locating specific passages in the course materials that clarify debated issues, asking for clarification of course concepts, and providing personal experiences to illustrate/explain course concepts.

Non productive comments included poorly answering a comprehension question, repeating another student's answer to a comprehension question, answering factual questions inaccurately, giving repetitious or off-topic testimony, offering off-the-subject jokes, and answering instructor questions by reading directly from the course materials. See Appendix D for definitions and examples of each comment category. Definitions of the participation levels were based on the time of day, setting, and class size. The sections in which the study occurred

were in the evening and were considered relatively small for a college course. Contributing two comments to a classroom discussion has been considered to be a low rate of participation in classrooms with a small class size (Howard et. al., 2002). Thus, the following participation levels were used in this study: non-responders (0 comments), occasional responders (1 to 3 comments a day), frequent responders (4 to 6 comments), and dominant responders (7 or more comments).

At the end of each discussion day, the primary observer placed students into categories based on what percentage of their total contribution to class discussion was considered productive that day. In order to be classified as an A-level responder, a students' contribution to class discussion had to be 99 to 100% productive. A B-level responder made 89 to 98% productive comments on that discussion day. C-level responder's comments were 68% to 88% productive and a D-F-level responder's comments were less than 68% productive. This method of assigning students to particular participation levels based on specific criteria was similar to the descriptions of participation levels offered by Bruss (2009). However, the current study more specifically operationalized the criteria used to categorize students by participation-quality level. Specifically, the criterion for a student to be in a particular group was based on quartile cutoffs of average percent of productive participation for combined sections in Unit 1 (i.e., baseline). In Unit 1 of combined sections, 22.2 % were A-level responders (n = 11), 26.5% were B-level responders (n = 12), 28.8% were C-level responders (n = 13), and 24.4% were D-level responders (n = 11). Slightly uneven groups were due to ties in scores at quartile cut-offs.

To self-record their participation, students purchased a set of specially designed record cards (see Figure 1). The students were instructed to write on the card a brief summary of each comment they posed in class on days 1 through 4 in each unit. The students turned in their

record card at the conclusion of the class on days 1 through 4 in each unit. Each comment was numbered on the record card and additional space was provided on the back of the card for students who made more than three comments.

The syllabus and the instructor defined student comments as voluntarily asking questions, responding to instructor questions, or offering additional perspectives on issues in class discussion. Choral responding was not considered a comment. A brief interchange between the instructor and the student in order to clarify the student's comment was considered one comment. In addition to recording their participation on each discussion day, students were required to provide the following information on their record cards the first 3 days of each unit: attendance, instructor notes homework completion, articles homework completion, and display of name card. Each of the daily-credit tasks was worth one point; thus, the student could earn a total of four points each discussion day for completion of daily-credit procedures. By requiring students to record additional information on their student record cards, the likelihood of students submitting their record cards at the conclusion of the class period should increase (Krohn et al., 2010). On the fourth day of each unit students recorded their comments only, inasmuch as there was no other credit producing activity on that day.

Students also were instructed to record the timeliness of their comments in the designated locations on the record card as either timely (T) or repetitious (R). A timely comment was defined by the syllabus and the instructor as a comment that pertained to the topic being discussed and not repeating earlier student comments or points already explained by the instructor. Repetitious comments included questions that had been previously asked by another student. Following a repetitious comment, the instructor provided feedback such as "that's a

point we already discussed," "that's the same question Jim asked," or "Sally explained that issue a few moments ago."

A secondary observer, also a GTA who previously taught four semesters of the course, recorded the quantity and quality of student comments on the third day in each unit to permit assessment of agreement between the two observers' records of the quantity and quality of student comments. The primary and secondary observer used the Student Discussion Form (see Appendix E) to record the quality and quantity of student comments. Students were listed by first name alphabetically so that raters could easily locate them on the form (Aspiranti, 2010; Krohn, 2010). Both observers practiced rating the quality of student comments from a sample script (see Appendix F) of a typical class discussion, a procedure that permitted comparisons between the ratings of the two observers and further refinement of the rating scale. The primary observer practiced rating on the first day of Unit 1 in a section of the course not included in the study. The secondary observer also practiced on the second and fourth day of Unit 1 in the sections of the course included in the study.

Participation Contingencies

Participation contingencies were designed primarily to increase student participation in selected units. Additionally, the contingencies allowed the researcher to investigate whether the contingencies balanced participation across students. On 2 randomly selected days in two units, the students were given credit for participating in class. The two units in which the students were to receive credit for participation were randomly selected and announced at the beginning of the course. The days on which students would receive credit were randomly selected in class on the last day of each credit unit. Students received three points for their first comment and two points for one additional comment. If a student earned a total of five points (i.e., made two

comments) on both days in each selected unit (i.e., 4 days total), the student received 10 additional points at the conclusion of the course. If a student earned points on both selected days in each unit, but did not earn maximum credit (i.e., 5 points) on all 4 days, the student received 5 additional points at the conclusion of the course. Thus, students could earn up to 30 points for their participation (10 points in each of two units and 10 additional points for achieving maximum credit in both units), constituting approximately 6% of the students' course grade. Foster et al. (2009), Krohn (2010), and Krohn et al. (2010) have demonstrated that crediting students for class participation equivalent to 3 to 5% of course credit is effective in increasing the percentage of initially low-responding students who subsequently volunteer comments during class discussion.

Student Participation Survey

A 50-item survey was used to assess students' history/confidence in their class participation, expectations for their participation in this course, and personal benefits of participating in class. Both Krohn (2010) and Aspiranti (2010) examined factors in the participation survey used in the current study. Krohn's three-factor model was used in the current study rather than Aspiranti's seven-factor model due to significant differences between participation levels for common factors in both studies, common significant predictors of participation levels, and a more parsimonious factor structure extracted by Krohn. The survey was posted on the course website and students were required to complete the survey at the beginning of the course in order to receive 5 points for the completion of the survey. Data were used to predict the level of quality of participation, as well as to clarify why some students participated minimally in class despite credit incentives. Each question was followed by five Likert-type response options.

Assessment of Critical Thinking

Students completed the 40 item *Watson-Glaser Critical Thinking Appraisal-Form S* (WGCTA) on the first day of class for bonus credit (approximately 1% of course credit). The WGCTA-Form S provides a measure of critical thinking in adults. Scores on the WGCTA-Form S have been correlated with various achievement measures (i.e., ACT, SAT, and reading achievement) and is considered a reliable measure (i.e., internal consistency and test reliability of .81) (Watson & Glaser, 1994). Data from the WGCTA were used to predict quality of participation levels among students.

Unit Exams

Students completed a 50-item multiple choice exam at the conclusion of each unit. Approximately two-thirds of the exam items required students to apply concepts outlined in course materials, many of which were discussed in depth in class discussion (Wallace & Williams, 2003). The remaining items required students to recall information found in the course materials. The unit exams constituted approximately 70% of the final course grade.

Participation Contingencies and Research Design

Each section received treatment aimed to increase the frequency of student comments in two of the five units. Both sections began with a period of baseline data collection in Unit 1. Following Unit 1, Section A received treatment while Section B received an extended baseline. The credit and non-credit units alternated for the remaining units. The assignment of the credit sequences to the two sections was based on random selection. See Table 1 for a visual display of the treatment conditions across the two sections. The nature of this research is considered a group withdrawal design.

Chapter III

Results

The results of this study are presented in the following order: (1) descriptive analyses of quality and quantity of participation by gender and academic classification; (2) inter-rater agreement on quality and quantity ratings of comments; (3) graphs of the percentage of students in each qualitative grade category; (4) proportions comparisons of the quality and quantity of comments between treatment and non-treatment phases; and (5) the potential of participation levels, participation survey factors, critical thinking scores, and exam score to predict inclusion in participation quality categories.

Descriptive and Correlation Analyses

Quality of participation. Table 2 presents the mean percent of productive comments contributed daily to class discussion for the entire course, credit units, and non-credit units based on gender and academic levels. Overall, the participation of females (M = 82.4, SD = 14.5) was significantly more productive than that of males (M = 62.4, SD = 28.0), *Welch's* $t(17) = 2.40$, p = .027. Furthermore, on average, females consistently contributed a greater percentage of productive comments during both credit and non-credit units in both Section A and B. However, males (n = 11) only constituted 25% of the total sample. Due to the under-representation of males and lack of homogeneity of variance among males and females in participation quality, no further attempt was made to control for gender or to make gender comparisons in participation. There was no difference in the average percent of productive comments made daily in the course overall, credit units, and non-credit units among academic years. Freshmen and graduate students were not adequately represented in this sample due to small ns (two and three,

respectively). In addition, students' GPA was not significantly related to the mean percent of productive comments contributed daily per students, $r = -.03$, $p = .913$.

Frequency of participation. Differences in the average number of comments students made throughout the course were investigated across gender (male and female) and academic levels (freshman, sophomore, junior, senior, graduate). Although females reported slightly more daily participation than males, they did not differ significantly in the daily average number of comments throughout the course, daily average number of comments during credit units, or daily average number of comments during non-credit units. Additionally, no significant differences in overall participation, participation during credit units, and participation during noncredit units were obtained across academic levels. Table 3 presents the mean number of comments students contributed for the entire course, credit units, and non-credit units. Additionally, students grade point average (GPA) was not significantly associated with the average number of comments students made throughout the course, $r = -.01$, $p = .955$.

Reliability

Quality of participation. The primary observer rated each comment in all discussion classes based on the rating system described in Appendix D, providing a tally of the types of comments made by each student each day. Additionally, the secondary observer provided a tally of the types of comments made by each student on the third day in each unit. The researcher compared the observers' data to assess the inter-rater agreement on participation quality.

Productive comments. The average inter-observer agreement (IOA) between raters was calculated for each individual quality category, yielding 15 reliability indicators on five occasions (i.e., one day per unit) throughout the course. The average IOA for productive comments is presented in Table 4. Overall, there was a 90% agreement between the primary and

secondary observers' records of the number of productive comments. If the primary observer identified a comment as comprehension and the secondary observer identified the same comment as factual, this combination would result in an agreement in the overall productive category, but it would be counted as a disagreement in the individual categories. Thus, overall IOA on productive comments was not an average of IOA among individual categories. Compared to least used categories, those that were used most frequently (i.e., 14 to 34 agreements on average per unit) resulted in higher reliability: Comprehension (80%), Testimony (63%), and Factual (18%). Raters generally agreed the least on productive comment categories that were rated seldom (i.e., 1 to 12 agreements on average per unit): Clarify (43%), Locate (33%), Other (33%), Answer (19%), and Skeptical (0%). The overall agreement for the number of productive comments contributed by each student in Section B (81 to 91% across units) was slightly lower than in Section A (89 to 98% across units). In addition, the percent of agreement for productive comment categories appeared to remain unaffected by the presence and absence of a credit unit. The ability of raters to categorize productive and nonproductive comments using the system described previously was also assessed with a Pearson-product-moment correlation (see Table 5). Correlations between the primary and secondary observers' records of the number of productive comments contributed by each student ranged from .96 to .98, which are considered to be high (Cohen, 1988). The correlations were similar across sections and units.

Nonproductive comments. The average IOAs between the primary and secondary rater for overall and specific nonproductive comment categories are presented in Table 6. Both raters used categories considered nonproductive on 108 occasions on the 5 inter-rater days (an average of 18 agreements per unit). These frequencies are smaller than those for the use of productive categories, which were used on 1,120 occasions by both raters on 5 inter-rater days (an average

of 35 agreements per unit). Thus, due to fewer occurrences, disagreements between raters more greatly affected the percent of agreement between the primary and secondary raters. The percent of agreement for nonproductive comments was 49 across sections and units, and the overall nonproductive categories ranged from 0 (Poor Testimony) to 57% (Poor Factual). The percent of agreement on nonproductive comments appeared to be unaffected by the presence or absence of a treatment condition. The percent of agreement between primary and secondary raters on total nonproductive comments made by each student was generally higher in Section B (25 to 81%; average 52%) than in Section A (42 to 65%; average 41%).

Quantity of participation. As described previously, a primary observer rated quality and quantity of student comments on each discussion day. Furthermore, a trained secondary observer rated quality and quantity of student comments on the third day of each unit. Data (i.e., the number of comments made by each student) recorded on the raters' form were compared to data from the students' record cards, which the students submitted at the end of each discussion day. Inter-rater agreement measurements were calculated for the following variable pairs: (1) frequency of student-recorded participation and rater-recorded participation and (2) frequency of participation recorded by primary and secondary raters.

Table 7 summarizes the reliability, measured by average inter-observer agreement (IOA) between the student-recorded participation frequency and primary rater's record. The average IOA was calculated in two steps. First, the percent of agreement between the student and primary rater was calculated by dividing the smaller tally by the larger tally from either the student or primary rater for each student in each category. Second, the average percent of agreement among students was computed for each category in each unit, yielding the average IOA for each category in each unit. As recommended by Krohn (2010), instances in which both

observers recorded no tallies of a particular category were excluded from the inter-observer agreement analyses in order to avoid inflating the reliability indicators.

The average inter-observer agreement between student-reported participation frequency and the frequency of participation reported by the primary observer, which was computed for each day in each unit for both sections, ranged from 35 to 92%, with an overall average of 64% per day. In Section A the average IOA per day ranged from 35 to 69% (average 58%) in noncredit units and from 64 to 79% (average 71%) in credit units. In Section B the average IOA per day ranged from 45 to 66% (average 58%) in noncredit units and from 62 to 92% (average 68%) in credit units. Student agreement with the primary observer was somewhat stronger in the treatment phases than the non-treatment phases. Analyses of student under and over reporting, described below and displayed in Appendix G, suggest that the discrepancy between student and observer ratings was largely due to student under-reporting during non-credit units.

Table 8 indicates the average inter-observer agreement of student-recorded participation frequency with the primary and secondary rater's records of participation, as well as the reliability of rater-recorded participation frequency on inter-rater days by section and unit. With both sections combined, the average IOA representing the degree of agreement of participation frequency between students and the primary observer on inter-rater day ranged from 56 to 67% (average 60%). Average inter-rater agreement between students and the secondary observer ranged from 50 to 71% (average 58%). Lastly, the overall inter-observer agreements between primary and secondary observers ranged from 88 to 97% (average 92%). Primary and secondary observer agreement on the number of comments made by students was generally stronger than observer-student agreement. As has been found previously, this pattern suggests that students

may be under and/or over-reporting the number of comments students contribute to class discussion (Aspiranti, 2010; Krohn, 2010; Krohn et al., 2010).

Appendix G displays an in-depth summary of student under- and over-reporting. The extent of under-reporting and over-reporting of the number of comments was assessed using a statistical analysis that examines the difference between two proportions (Ferguson & Takane, 1989). Proportions were based on the number of students present each day and the number of students who over-reported and under-reported their comments each day. Students who recorded that they contributed more than the number indicated by the primary observer for that student were considered to over-report their participation. Students who recorded less comments than reported by the primary observer were considered to under-report their participation. There were 4 days in each credit and non-credit unit in which both the student and primary observer recorded the number of comments contributed by each student. Thus, it was possible to compare student and primary observer records across 4 days in each credit and non-credit unit.

The percentage of students who under-reported the number of comments they contributed to class discussion in non-credit units ranged from 34.2 to 64.9 (average 50%). The proportion used to represent the percent of students under-reporting their participation in non-credit units was 146/295, where 146 represents the sum of students who under-reported their participation across the 8 days in non-credit units and 295 represents the sum of students present across the 8 days in non-credit units. Most under-reporting in non-credit units occurred when students contributed comments beyond the participation treatment-level (i.e., 2 comments). Of the instances in which students under-reported their participation, 18.5% (i.e., 27 out of 146 instances) would have resulted in students' not receiving credit had credit been awarded in non-credit units. During credit units, the percentage of students who under-reported their

participation ranged from 20.5 to 68.3% (average of 49.3%). Across the 8 days in combined credit units, a total of 149 out of a total of 308 students under-reported the amount of their participation. Only 3.4% (i.e., five out of 149 instances) of students who under-reported their participation in credit units failed to receive due credit; thus, under-reporting below the credit level was much less prevalent in credit than non-credit units. Students did not under-report their participation significantly more in non-credit units (146/295) than in credit units (149/308), p = .392).

Only 1 to 2 students per day recorded more participation than the primary observer during non-credit units; thus, over-reporting of participation in non-credit units was rare. Across the 8 days in non-credit units, students who over-reported their participation ranged from 0 to 7.9% (average 3.7%). The total number of students who over-reported the amount of their participation on the 8 non-credit days was 11 out of a total of 295 cumulative students present. Thus, the proportion used to represent the percent of students over-reporting participation during non-credit units was 11/295. On the other hand, students who over-reported their participation in credit units ranged from 5 to 15.4% (average 7.5%). The total number of students who over-reported the amount of their participation across the 8 days in non-credit units was 23 out of a total of 308 students present. The proportion of students over-reporting in credit units (23/308) was significantly greater than the proportion of students over-reporting during non-credit units (11/295), p = .023. The majority of students who over reported their comments during credit units, reported beyond treatment-level participation (i.e., recorded more than two comments when the observers recorded only two comments). However, 43.5% (i.e., two to three students per day) of students who over-reported their participation received undue credit. Most students who undeservedly received credit (87.5%) over-reported their comments to receive partial credit.

There was only one instance in which a student over-reported their participation in order to receive full credit. Of the 45 students who self-recorded their participation during credit units, 10 students earned credit that was not equivalent to the credit that should have been awarded according to the records of the primary observer. The majority of these students (n = 8) over-reported their participation to receive credit on only 1 out of the 8 days credit for their participation could be earned. Furthermore, only two students were found to consistently over-report their participation for credit. One student over-reported his participation for partial credit on 3 out of 8 days and another student over-reported his participation for partial credit on 2 out of 8 days in the credit units.

The Effect of Credit Contingencies on Participation

Equivalence of baseline course knowledge. Students' prior knowledge about course concepts could potentially affect the quality and quantity of student participation. In order to assess whether Section and A and B had equivalent initial course knowledge, Unit 1 exam scores were compared across sections using a one-way analysis of variance. Students typically had some prior experience with Unit 1 concepts (i.e., physical fitness, dietary knowledge); thus, it was the best pre-course knowledge measure available. There was not a significant difference in Unit 1 exam scores between Section A (M = 39.58, SD = 7.06) and Section B (M = 36.91, SD = 5.26), $F(1, 45) = 2.09$, $p = .156$. These results provide evidence for equivalent course knowledge across sections.

Equivalence of baseline participation. Previous studies (Foster et al., 2009; Krohn et al., 2010) have noted that the inflated levels of participation at the onset of previous semesters of the same course were likely due to concept familiarity of Unit 1 and/or novelty of the format of the course. In the current study, the baseline was extended to the second Unit in Section B.

Unlike previous studies, there is no evidence for inflated Unit 1 participation quantity in the current study. The mean number of comments per day per student increased slightly in Section B from 4.28 in Unit 1 (baseline) to 4.71 in Unit 2 (extended baseline/non-credit unit). In Section A, the mean number of comments per day per student decreased slightly from 4.26 in Unit 1 (baseline) to 4.22 in Unit 2 (credit unit). It is important to note that the major purpose of implementing a credit contingency for participation quantity was to balance participation among class members rather than to increase overall participation. There was minimal evidence of an inflated baseline for participation quality in this study. The average percent of productive comments per day per student decreased from 84% in Unit 1 (baseline) to 82% in Unit 2 (extended baseline/non-credit unit) in Section B. In section A, the average percent of productive comments per day per student increased from 78% in Unit 1 (baseline) to 83% in Unit 2 (credit unit).

One-way analyses of variance were used to assess differences in students' baseline participation quality and quantity between Section A and B. The analysis investigating section differences in students' average percent productive comments per day in Unit 1 was not significant, $F(1, 44) = .98$, $p = .327$. Students in Section A ($M = 78\%$, $SD = 22\%$) and Section B ($M = 84\%$, $SD = 15\%$) were equally productive. In addition, the analysis investigating section differences in students' average number of comments contributed per day in Unit 1 was not significant, $F(1, 44) = .01$, $p = .985$. On average, students in Section A ($M = 4.26$, $SD = 3.10$) contributed the same number of comments per day as students in Section B ($M = 4.28$, $SD = 3.15$) during Unit 1.

Given that productive comments were more frequently observed in each discussion day throughout the course than non-productive comments, the distribution of average percent of

productive comments made each discussion day was skewed such that 80% of students fell between 56 to 100% productive. Students' qualitative level was not based solely on the total number of productive comments due to the possibility that students could make several productive comments, yet contribute an equal number of non-productive comments. Thus, percent of productive comments is a more accurate portrayal of a student's quality of discussion on any given discussion day. The percent of productive comments was calculated by dividing the number of productive comments contributed by the total number of comments contributed per day. Non-responders were not included in any participation quality level given that they did not comment during class, thus quality of their participation could not be assessed.

Effect of Credit on Quality of Participation. Visual representations of the percent of students at each qualitative level each day for each unit within each section are displayed in Figure 2. In Section A, the number of students classified as D-level students decreased by an average of 4% in credit units. Overall, C-level students enrolled in Section A increased by 6% in credit units and B-level students remained the same. A-level students increased by an average of 4% in credit units in Section A. Section B displayed a similar, but more consistent trend. The levels in Section B generally decreased slightly in credit units: D-level (1%), C-level (3%), B-level (.33%). A-level students increased by 11% in credit units in Section B.

A statistical proportions test was used (Ferguson & Takane, 1989) to assess the magnitude of the difference in the percent of students participating at each qualitative level in credit and noncredit units for each section. Initially, the average percent of students at each level was calculated for all non-credit units (i.e., 2 out of the 5 units) and credit units (i.e., 2 of 5 units) for each section. These two proportions were compared using the proportions test previously described. The average percent of students at each qualitative level across units was also

compared for treatment pair 1 (Unit 2 and Unit 3) and treatment pair 2 (Unit 4 and Unit 5). These proportions are displayed in Table 9. There were no significant differences in the percent of students at each level between credit and non-credit units, within treatment pair 1 or treatment pair 2. Therefore, the proportion of participation students at each qualitative category remained fairly constant across the units and was not affected by the credit contingency.

Another way to discern whether quality of discussion was affected by the credit contingency is to consider the discussion as a whole on each day in each section. Figure 5 illustrates the percent of the discussion each day that was considered productive across units for both sections. Visual inspection reveals little treatment effect for the credit-contingencies. On average, productive discussion increased by 1.7% in credit units in Section A. In Section B, the 4.2% increase in productive discussion during credit units was also minor.

Table 10 displays the proportions and significance levels across sections for the percent of productive discussion between overall credit and non-credit units as well, as treatment pairs 1 and 2. There was a significant decrease in productive discussion in both sections during Unit 2, a credit unit for Section A and a non-credit unit for Section B, suggesting that the decrease in productive discussion was probably attributable to the difficulty of the unit rather than the credit contingency. Unit 2 is the unit in which students generally perform most poorly in the course. This pattern is similar to the pattern that Krohn (2010) described in her analyses of comment types (i.e., comments, questions, and anecdotes) across units. There was also a significant increase in Section B's quality of discussion in credit units; however, the non-credit units in this section are likely artificially low given that they included Unit 2. Therefore, we cannot say with certainty that Section B's increase in quality of discussion in credit units was due to the credit contingency alone.

Effect of Credit on Quantity of Participation. Figures 6 to 9 display the percent of students in each level of participation in each day, illustrating trends in frequency of participation across units (i.e., alternating treatment and non-treatment phases) in each section. Visual analyses of the percentage of students in each participation level reveals little evidence of a significant treatment effect on non-, occasional, frequent, and dominant responders. Non-responders increased slightly in non-credit units by an average of 10% in Section A and 8% in Section B. In Section A, occasional responders (1 to 3 comments a day) increased by an average of 2% in non-credit units. On the other hand, occasional responders in Section B decreased by 8% in non-credit units. Changes among frequent responders were inconsequential: 1% increase in non-credit units in Section A and 2% decrease in non-credit units in Section B. Overall, dominant responders decreased by 8% in non-credit units in Section A. Conversely, dominant responders increased by an average 2% in non-credit units in Section B.

In order to assess whether the frequency of participation increased significantly in credit units, the researcher used the proportions test described above (Ferguson & Takane, 1989). The average percent of students at each level in credit units were compared to the average percent of students at each level in non-credit units in each section (see Table 11). This analysis revealed no significant differences between the percent of students in credit and non-credit units across all participant levels (i.e., non-responder, occasional, frequent, dominant). Additionally, when the proportion of students at each level were compared within treatment pair 1 and treatment pair 2 in each section, there were no significant differences across all participation levels across units within sections. In sum, the credit contingency appeared to have little effect on the amount of participation. This finding is inconsistent with previously reported treatment effects from participation credit (Foster et. al., 2009; Krohn, 2010). It may be that the credit contingency (i.e.,

credit for 1 to 2 comments a day) was too low in order to balance class participation, given the increased opportunity for participation in classes with about 25 students.

Intra-Student Comparisons of Frequency of Comments

One purpose of imposing a credit contingency for frequency of student commenting is to balance student participation. In order to assess whether participation was more balanced in credit units, the researcher tracked initially low, medium, and high responders' participation across phases within sections. Students were classified in these categories based on the quartile cut off scores for the average number of comments made in Unit 1. Students who contributed 2 comments a day or less (bottom 27%) were identified as low responders. Students who contributed 3 to 5 comments a day in Unit 1 were identified as Medium Responders (middle 45%), and students who offered 6 or more comments a day (top 28%) in Unit 1 were classified as high responders.

A total of 12 students (*ns* = 5 and 7 for sections A and B) were identified as initially low responders (i.e., in Unit 1). None of the low responders were considered non-responders in that all students in the study participated on at least one occasion in Unit 1. Figure 10 displays the percent of initially low-responding students who participated in subsequent days throughout the course. In Section A, the average percent of low responders who participated in credit units (43%) was about the same as the average percent of low responders who participated in noncredit units (44%). On the other hand in Section B, the average percent of low responders who participated subsequently decreased from 73% in credit units to 60% in noncredit units.

Proportions testing (Ferguson & Takane, 1989) was used to assess significant difference between credit and noncredit units (baseline included in non-credit units), treatment pair 1, and treatment pair 2 for the average number of initially low responders who subsequently

participated each day per unit. The average proportion of low-responding students who participated in credit units (2/5) was not significantly different than those who participated in noncredit units (2/5). Similarly, there was no significant difference in the average proportion of low-responding students who subsequently participated in credit units (4/6) and noncredit units (4/6). Differences within treatment pair 1 and 2 were also not significant for both sections.

Table 12 displays how many initially low-responding students fell into low, medium, and high categories in subsequent units, illustrating another way to investigate the participation of initially low responders throughout the course. The data in Table 12 differ from Figure 10 in that it tracks initially low responders by unit rather than by day. With the exception of the first credit unit in Section B, the majority of initially low-responding students remained low responders in the two subsequent credit units in both sections: 100% and 80% in the first and second credit unit respectively for Section A and 43% and 57% in Section B.

Table 13 displays the number of Medium-Responding students in Unit 1 who fell into low, medium, and high categories in the subsequent units. The movement of Medium-Responders in credit and noncredit units was inconsistent. In the two credit units of Section A, 13% of initially Medium Responders in the first credit unit and 33% of initially Medium Responders in the second credit unit moved to the low category. In Section B, 0% and 28% of initially Medium Responders in the first and second credit unit respectively in Section B moved to the low category. Additionally 15% and 33% in the first and second credit units of Section A and 71% and 0% in the first and second credit units of Section B were subsequently classified as high responders in credit units. In the two noncredit units, 38% and 62% of initially medium responders in the first and second noncredit units in Section A, and 31 to 33% in Section B moved to the low category. Thirty-one percent and 15% in the first and second noncredit unit in

Section A and 14% and 29% in the first and second noncredit unit in Section B moved to the high category.

Table 14 displays the number of initially high-responding students who fell into low, medium, and high categories in subsequent units. Few initially high-responding students moved from their initial classification group. If these students moved, they were more likely to move to the medium category than the low category. In Section B, 14% (n = 1) of initially high responders moved to the low category in the first non-credit unit. Otherwise, in all units, regardless of the presence of a credit contingency, one to three initially high responders moved to the Medium category in both sections.

Predicting Consistent Inclusion in a Qualitative Category

A stepwise discriminant-function analysis was conducted in which eight course variables were considered potential predictors of membership in two participation groups, high- and low-quality responders. A stepwise discriminant-function analysis is useful in determining whether some predictors are considered higher priority than others in predicting classification into groups. It also can reduce significant multicollinearity among independent variables and produce a parsimonious model. Personal History and Confidence Regarding Participation, Expectation for Discussion in College Classes, and Personal Benefits of Participation, survey score of combined items constituting the three factors, and total survey score served as possible predictors of high- and low-quality participation. In addition, critical thinking scores, average exam score, and average daily course participation were also examined as predictors of the high and low levels of participation quality. Participants were classified into low- and high-quality groups based on the quartile cut-off scores for their average percent productive scores. Students with low-quality participation (n = 11) were only 68% productive or less on average per day.

Students with high-quality participation (n = 10) were 90% or more productive on average each day. As demonstrated by Krohn et al. (2010), participants in the medium participation group are very variable in their participation and course performance, as well as smaller in number than low and high groups within units. Thus, in order to most accurately examine differences between students with high- and low-quality responders, the medium group was excluded. Also, given that there were no significant treatment effects for the quality of participation and the small number of participants in high- and low-quality groups, the sections were combined and not separated for all prediction analyses. One product of a discriminant-function analysis is an examination of significant mean differences in each independent variable between low- and high-quality groups via univariate analyses of variance. Mean differences in the course predictors between high- and low-quality groups generally exist when there is a significant predictor.

Participation factors and participation quality. The participation survey used in this study was also used in the Krohn (2010) study. Krohn identified three factors within the survey: Personal History and Confidence Regarding Participation, Expectation for Discussion in College Classes, and Personal Benefits of Participation. Chronbach's alpha was used to assess the internal consistency for the total survey (.88), combined factors (.90), History/Confidence factor (.83), Expectation factor (.86), and Personal Benefits factor (.67). With the exception of the Personal Benefits factor, all internal consistency measures are considered acceptable based on the social sciences research standard of .70 (Garson, 2008).

Table 15 displays means and standard deviations for low- and high-quality responders for the total survey, combined factors, History/Confidence factor, Expectation factor, and Personal Benefits factor. A series of discriminant-function analyses was conducted in order to determine

whether significant predictors collapsed into other variables once all variables were included as predictors in a discriminant-function analysis. First a discriminant-function analysis was conducted for each of the following variables: total survey, combined factors, History/Confidence factor, Expectation factor, and Personal Benefits factor. Then, a discriminant-function analysis was conducted that included all of the survey variables listed previously as predictors. Students' total scores on all items included in the participation survey was not a significant predictor of high- (M = 184.70, SD = 13.34) and low- (M = 171.55, SD = 16.72) quality responders, X^2 (1) = 3.47, p = .063. A student's total score on all items included in the three factors was a significant predictor of high- (M = 85.70, SD = 10.83) and low- (M = 74.81, SD = 12.46) quality participation, X^2 (1) = 3.95, p = .047. Score patterns on the combined factor items also produced the most reliable measure among the participation survey measures for this particular sample. A student's score on the History/Confidence factor was an indicator of students' previous experience with class participation and confidence in their ability to speak in class such that higher scores indicated a greater history of participation and confidence in contributing to class discussion. Students' scores on the History/Confidence factor was a significant predictor of high- (M = 30.50, SD = 4.33) and low- (M = 24.91, SD = 5.67) quality responders, X^2 (1) = 5.34, p = .021. Neither the Expectation factor (p = .226) nor the Personal Benefits factor (p = .120) were a significant predictor of high- and low-quality groups. When the total survey, combined factors, History/Confidence factor, Expectation factor, and Personal Benefits factor were considered as independent variables in a stepwise discriminant-function analysis, only the History/Confidence factor was considered a significant predictor of high and low-quality responders, X^2 (1) = 5.34, p = .021.

Critical thinking and participation quality groups. Students' could score a maximum of 40 points on the *Watson-Glaser Critical Thinking Appraisal-From S* (WGCTA-S). A discriminant-function analysis was conducted in order to determine whether low- and high-quality responders were distinguishable by their critical thinking skills alone. The analysis revealed that critical thinking was not a significant predictor of low- (M = 24.73, percentile rank = 5th, SD = 4.84) and high- (M = 27.50, percentile rank = 20th, SD = 7.28) quality groups, p = .313.

Exam scores and participation quality groups. As described earlier, students completed 50 item multiple-choice exams at the conclusion of each of the five course units. Of particular interest is exam performance among the qualitative levels. The means and standard deviations of exam scores for the qualitative levels are presented in Table 16. The distinction of high- and low-quality groups described previously was used to compare average exam scores (i.e., mean of all five exam scores) among qualitative levels. Similarly, students were classified into low- and high-quality levels based on their average percent of productive comments within each unit. Thus, a student may belong to the low-quality group in Unit 1 and the high-quality group in Unit 2. Students' classification based on a particular unit's productive participation average was used to compare that unit's corresponding exam scores. A series of discriminant-function analyses were used to compare exam score means among the qualitative groups in each unit and overall (i.e., a total of six discriminant-function analyses). A students' average exam score was a significant predictor of overall high- (M = 42.08, SD = 3.92) and low- (M = 37.52, SD = 5.01) quality groups, $X^2(1) = 5.37$, p = .033. Students' could score a total of 50 points on exams; low-quality responders averaged 75% correct on exams and high quality responders averaged 84% correct on exams. Students' performance on the Unit 1 exam was a significant

predictor of high- (M = 40.80, SD = 4.48) and low- (M = 36.11, SD = 4.96) quality responders, $X^2(1) = 5.62$, $p = .018$. Unit 4 exam scores also significantly predicted student membership in the unit's corresponding high- (M = 39.91, SD = 5.41) and low- (M = 34.55, SD = 9.00) quality groups, $X^2(1) = 4.21$, $p = .040$. Scores on the Units 2 ($p = .067$), 3 ($p = .290$), and 5 ($p = .873$) exams were not significant predictors of participation-quality groups.

Frequency of participation and qualitative levels. The mean number of comments contributed to class per day throughout the class and each unit for corresponding low- and high-quality responders is presented in Table 17 and Figure 11. The average number of comments contributed to class discussion on a daily basis was a significant predictor of high- (M = 5.44, SD = 3.15) and low-quality responders (M = 1.07, SD = .78), $X^2(1) = 13.27$, $p < .001$. High-quality participating students also contributed significantly more comments on average than low-quality participating students in all units: Unit 1, $X^2(1) = 5.28$, $p = .022$; Unit 2, $X^2(1) = 8.47$, $p = .004$; Unit 3, $X^2(1) = 12.22$, $p < .001$; Unit 4, $X^2(1) = 7.40$, $p = .007$; and Unit 5, $X^2(1) = 5.00$, $p = .025$.

Course variables' prediction of qualitative groups. In order to determine which course variables are most useful in distinguishing between high- and low-quality groups, all course variables (i.e., History/Confidence factor, Expectation factor, Personal Benefits factor, critical thinking scores, average exam scores, and average number of comments contributed each day) were included as predictors in a discriminant-function analysis with high- and low-quality groups as the criterion variable. Due to significant mulitcolinearity (i.e., correlations greater than .70), between the combined factor items and each of the three survey factors, as well as between the total survey score and each of the three survey factors, combined factor items and total survey score were removed from the analysis. Thus, the History/Confidence factor, Expectation factor, Personal Benefits factor, critical thinking scores, average exam scores, and average

number of comments contributed each day were included as predictor variables. Table 18 summarizes the discriminant-function analysis results. One significant discriminant function was calculated, $X^2(2) = 18.62$, $p < .001$, Canonical $R^2 = .80$. Thus, the function accounted for about 80% of the total relationship between the course variables and qualitative groups. Critical thinking scores and average daily frequency of participation were identified as significant predictors of low- and high-quality responders. Although average exam scores were a significant predictor of high- and low- quality participation when included as the only predictor and critical thinking was not a significant predictor of high- and low-quality groups when included alone, the average number of daily comments in conjunction with the critical thinking scores explained the most variance among high- and low-quality responders. Average daily frequency of participation was a stronger predictor of qualitative groups than critical thinking scores. For every one-unit increase in the number of comments contributed on average each day, a students' discriminant score increased by 1.14 standard deviations above the mean. In a discriminant-function analysis, the closer a students' discriminant score is to zero, the less likely the predictors indentified in the factor are able to accurately identify the student as a low- or high-quality responder. Thus, the more comments that a student contributed on a daily basis, the more likely the student was considered a high-quality responder. Additionally, for every one-unit increase in critical thinking scores, a student's discriminant score increased by .75 standard deviations above the mean. This indicates that students with higher critical thinking scores are more likely to be classified as high-quality than low-quality responders. Together, average daily frequency of participation and critical thinking scores accurately predicted 95.2% of the cases. In fact, the two course variables accurately predicted membership to low-quality responders with 100% (11 of 11 students) accuracy and to high-quality responders with 90% (9 of 10 students) accuracy. Given

that we used the same subjects used to create the model, the percent correctly classified is likely an overestimation of how well the model performs on other subjects. Therefore, a cross validation procedure was completed in which each case was classified by the functions derived from all cases other than that case. This method produced the exact same prediction accuracy as the original cases, providing evidence for the generalizability of the model.

Differences among Qualitative Levels within Each Unit

In order to investigate how the number of comments a student contributed to class discussion was related to the quality of their participation, a proportions test was used to determine the differences in proportionality of the average number of students who were non-responders (0 average comments per day), occasional responders (1-3 average comments per day), frequent responders (4-6 average comments per day), and dominant responders (7 or more average comments per day) at each qualitative level of participation (A-level, B-level, C-level, and D-level) for each unit. It is important to note that a student's qualititative level was based on the average percent of productivity throughout the course. If a student did not participate or was absent on a given day, that day was not included in the computation of the average percent productivity. Therefore, nonparticipation was not a factor included in assigning students to qualitative levels. Furthermore, a student considered an A-level responder could theoretically also be considered a non-responder on any given discussion day as long as the student's participation on days that the student did contribute to class discussion was highly productive. This analysis provides a more in-depth investigation of the participation patterns within qualitative levels than the discriminant-function analyses because all qualitative levels are included and participation patterns are tracked across units and within sections. The proportions

and significant differences between pairs of participation levels within qualitative levels are displayed in Table 19.

Overall, within the A-level group, there were significantly more occasional responders (6/10) than non-responders (0/10, $p = .002$); significantly more frequent responders (3/10) than non-responders (0/10, $p = .030$); and significantly more occasional responders (6/10) than dominant responders (1/10). Thus, in general, A-level students were most likely to be occasional responders. Among the A-level responders in Unit 1, there were significantly more occasional responders (5/9) than non-responders (0/9, $p = .004$). In Unit 2, there were significantly more occasional responders (7/10) than non-responders (0/10, $p = .001$) and dominant responders (0/10, $p = .001$) among A-level responders. Also in Unit 2, there were more frequent responders (3/10) than non-responders (0/10, $p = .03$) and dominant responders (0/10, $p = .03$) among A-level responders. In Unit 3, there were significantly more occasional responders (7/12) than non-responders (0/12, $p = .001$), frequent responders (3/12, $p = .049$), and dominant responders (2/12, $p = .018$) among A-level responders. Furthermore, there were significantly more frequent repsonders (3/12) than non-responders (0/12) among A-level responders in Unit 3, $p = .032$. In Unit 4, there were significantly more occasional responders (7/9) than non-responders (0/9, $p = .001$); frequent responders (2/9, $p = .009$); and dominant responders (0/9, $p = .001$) among A-level responders. In the final unit of the course, there were significantly more occasional (3/6) and frequent responders (3/6) than non-responders (0/6, $p = .023$) and dominant responders (0/6, $p = .023$). The significant trend in the final unit of the course highlights the general trend throughout the course for A-level students in that the students who made high-quality comments typically did not include non-responders or dominant students; rather, their frequency of commenting tended to be moderate (i.e., 1-6 comments a day).

There were significantly more frequent responders (3/9) than non-responders (0/9, p = .029), as well as significantly more dominant responders (4/9) than non-responders (0/9, p = .012) on average within B-level students. This indicates that B-level students were likely to be frequent or dominant responders and unlikely to be non-responders. Inspection of B-level students within individual units revealed that there were significantly more dominant responders (4/9) than non-responders (0/9) in Unit 1, p = .012. In Unit 2, there were also significantly more occasional (3/9), frequent (3/9), and dominant (3/9) than non-responders (0/9) among B-level responders, p = .029. There were significantly more frequent responders (3/10, p = .03) and dominant responders (5/10, p = .005) than non-responders (0/10) among B-level responders in Unit 3. In Unit 4, there were significantly more dominant students (6/10) than non-responders (0/10) and occasional responders (1/10, p = .03), as well as significantly more frequent responders (3/10) than non-responders (0/10) among B-level responders, p = .002. In general, B-level students typically tended to be dominant responders (i.e., commented on average seven or more occasions per discussion day) and participated at some level every day.

The only significant pair comparison among the C-level students was between non-responders (1/11) and occasional responders (5/11, p = .028.) Also, significantly more occasional responders (7/14) than non-responders (1/14, p = .006) emerged in Unit 1. There were significantly more occasional (5/11, p = .005), and dominant (4/11, p = .014) responders than non-responders (0/11) among C-level responders. C-level students' trend in comment frequency appears to be spread evenly over occasional, frequent, and dominant responders.

On average, there were no significant differences among participation level pairs within the D-qualitative level. In Unit 2, there were significantly more non-responders (3/8) than frequent responders (0/8) among D-level students, p = .027. In Unit 3, there were significantly

more non-responders (4/8) than frequent (0/8) and dominant (0/8) responders among D-level responders, $p = .01$. In Unit 4, there were significantly more non-responders (5/12) than frequent responders (1/12) among D-level responders, $p = .03$. Although not a consistent trend, D-level responders tended to be non-responders. It is important to note that a students' qualitative level was based on the average percent of productive comments contributed per day in a particular unit. The average percent of productive comments contributed per day included only days in which the student participated. Thus, if a student was present and did not contribute any comments on day 1 and that student's percent of productive comments was 100% on day 2, 50% on day 3, 50% on day 4, that student's average percent of productive comments would be 67% (i.e., 200/3). Therefore, it appears that D-level students likely participated infrequently and contributed poor comments when they did participate in class discussion.

Chapter IV

Discussion

The primary purpose of the current research study was to develop a subjective, systematic, and reliable measure of quality of student participation. The primary measure of reliability was inter-observer agreement. Inter-observer agreement between primary and secondary observers was computed for each of the 15 quality categories, the total number of productive comments contributed, and the total number of nonproductive comments contributed by each student. Additionally, the relationship between quality of participation and the following course variables were examined: average number of daily comments, average exam scores, critical thinking, personal history and confidence regarding participation, expectation for discussion in college classes, personal benefits of participation, combined participation survey factors, and total survey score. Investigating these relationships could help identify important course variables to target when attempting to increase the quality of students' class discussion.

Secondary research objectives include examining the effects of a random credit contingency on quality and quantity of participation. This investigation allowed the researcher to determine whether quality is affected when frequency of daily participation is rewarded. Additionally, the reliability of student self-recorded participation and the relationships between participation constructs and demographic variables were examined.

Reliability of the Participation Quality Measure

In order to measure participation quality, the researcher classified each student comment on 20 discussion days into one of eight productive categories or one of seven nonproductive comment categories. A secondary observer also rated student comments on 4 of the 20 discussion days. The major indicator for reliability of this measure was inter-observer agreement

(IOA). Also, Pearson product-moment correlations were computed in order to assess the reliability of the observers' ability to reliably rate the number of productive and nonproductive comments contributed by each student. According to Kazdin (1977) agreement between observers that reach 70 or 80% are often considered acceptable; however, he cautions that a prescribed satisfactory level of agreement may be challenging due to disagreement regarding methods of computing inter-observer agreement and the use of proportions to evaluate inter-observer agreement. Kazdin also notes that many variables should be investigated when judging whether inter-observer agreement is adequate, including variability and frequency of observed behaviors and the number of different categories included. Nonetheless, when inter-observer agreement falls within the widely acceptable range, researchers consider it appropriate to conduct further analyses using the behavior construct (Kratochwill & Wetzel, 1977).

Productive comment categories. In summary, the range of inter-observer agreement across units in combined sections are reported for the following productive categories: Comprehension (71 to 86%), Answer (9 to 30%), Factual (50 to 71%), Skeptical (0%), Locate (0 to 100%), Clarify (25 to 90%), Testimony (39 to 73%), and Other Productive (0 to 50%). Thus, IOA was highest and least variable in the category of Comprehension, which includes responses to instructor-posed comprehension questions and comments that relate, compare/contrast, or apply course concepts. Additionally, the IOA computed for the category Testimony (i.e., comments in which a student describes an experience to illustrate course concepts) was considered adequate in two of the five units (Kazdin, 1977). Inter-observer agreement reached adequate agreement in one out of five units in Factual and Clarify categories. Otherwise, inter-observer agreement was not considered adequate for the following variables across units in combined sections: Answer, Skeptical, Locate, and Other Productive.

The use of inter-observer agreement as a measure of reliability has been criticized due to its failure to consider agreement due to chance (House, Farber, & Nier, 1983) and its susceptibility to the frequency of behaviors measured (Bryington, Palmer, & Watkins, 2002). The percent of agreement between two observers regarding the occurrences of behaviors is highly influenced by the number of times the behavior occurs. For example, one disagreement on the classification of a behavior that was observed to occur at least twice according to one observer results in an IOA of 50%. On the other hand, one disagreement of a behavior that was observed to occur at least 10 times according to one observer would result in an IOA of 90%. Therefore, the agreement between the primary and secondary observers regarding the classification of Answer, Skeptical, Locate, and Other Productive categories was likely diminished by the infrequent number of occurrences reported by an observer. Comments classified in these categories were observed to occur least frequently among the eight categories, on average, each unit: Answer (n = 12); Skeptical (n = 2); Locate (n = 1); and Other Productive (n = 2). Comprehension was the category most often recorded by the observers (i.e., an average of 34 occurrences each unit). The average number of occurrences was 18 for Factual, 14 for Testimony, and 5 for Clarify.

The degree of agreement between the primary and secondary observer in classifying a comment as productive ranged from 84 to 95% across units in combined sections. This level of inter-observer agreement is well above the criterion to be considered adequate. It appears that the observers were more likely to agree when classifying a comment as productive than when classifying a productive comment into one of the eight productive comment categories. This suggests that the observers were likely to agree on a comment deemed productive; however, they may disagree as to which productive category the comment belonged. Therefore,

misclassifications of productive comments were likely across the eight productive categories. Furthermore, correlations between observers as to the number of productive comments rated for each student across units ranged from .96 to .98 in combined sections, indicating a high degree of association between the observers' records (Cohen, 1988). All correlations between observers and frequency of productive comments made by each student were significant at the $p < .001$ level. Given that the number of productive comments recorded by observers was the most reliable participation quality measure, the number of productive comments was used in the computation of the percent of productive comments made by each student, which was the primary measure of participation quality in this study.

Nonproductive comments. The IOA for nonproductive categories in combined sections ranged from 15 to 58% for Poor Comprehension, 31 to 100% for Poor Factual, 0% or not rated for Poor Testimony, 30 to 100% for Poor Joke, 0 to 100% for Reading Answers, 0 to 100% for Disagreeing Negatively, and 0% or not rated for Other Nonproductive Comments. Inter-observer agreement was highest among Poor Factual comments. In fact, in three out of five units, the observers achieved satisfactory levels of inter-observer agreement (i.e., above 70%) for poor factual comments. In two out of five units, the inter-observer agreement of the Poor Joke comment category was 100%. The Poor Comprehension comment category was the most widely used of the nonproductive comment categories; however, it was only recorded an average nine times per unit. Only one observer recorded poor Testimony, Disagreeing Negatively, and Other Nonproductive comments on one or two occasions throughout the entire semester. The average use of nonproductive categories by any one observer ranged from 1 to 12 for each unit. Thus, the nonproductive comments were used infrequently on a daily basis. As previously noted,

infrequent occurrence of nonproductive comments likely reduced the chance of achieving adequate inter-observer agreement in those categories (Bryington et al., 2002).

Inter-observer agreement between the number of nonproductive comments contributed by each student ranged from 46 to 74% across units in the combined sections. Only in the last unit did the observers achieve adequate agreement. In general, inter-observer agreement for nonproductive comments was higher in Section B than in Section A. In fact, the inter-observer agreement of nonproductive comments was higher in all but Unit 1 in Section B. Given that there was a similar number of occurrences of nonproductive comments across sections, the nature of the nonproductive comments in Section B were likely more distinguishable by the observers than those in Section A. In addition to lower inter-observer agreement, the correlations were also lower than those computed for productive comments (i.e., .62 to .86 in combined sections). The correlations between observers for the number of nonproductive comments rated for each student were significant at the $p < .05$ level.

Factors affecting inter-observer agreement. Kazdin (1977) highlighted several factors that may influence the degree of inter-observer agreement: 1) participant reactivity, 2) tendency of observers to recreate definitions, 3) complexity of the coding system, and 4) observer expectancies. Richards, Taylor, Ramasamy, and Richards (1999) also identify reactivity and "observer drift" as factors influencing inter-observer agreement (p. 91). The probability that participants experienced reactivity to the observers is likely low. Although the primary observer was seated at the front of the classroom, she was present for 80% of the class sessions. Thus, the students enrolled in the course had likely acclimated to her presence in the classroom. Furthermore, students were not aware that the observers were rating participation quality specifically. Instead the students were informed that the observers were investigating classroom

procedures. Also, students were not rating their own participation quality and they did not receive credit based on participation quality. Thus, even if they had been aware that the observers were rating their participation quality, the students were not inclined to alter their participation quality because of the observer's presence.

Furthermore, repeated measures ANOVAs revealed no significant differences in each section between inter-observer day and other days within each unit in average daily percent productivity per person and average number of comments per person. It appears that participants' participation patterns did not change significantly as a result of observer presence. Although, observer reactivity may have been problematic in the current study. Reid (1970) demonstrated that observers rated behaviors with a greater degree of accuracy when they believed their ratings were checked than when they did not. In the current study, it is possible that the primary rater classified student comments with more accuracy on the days in which the secondary observer was present (i.e., 25% of the total class discussion sessions) than on days in which the secondary observer was not present.

"Observer drift" occurs when observers begin to develop their personal interpretation of behavior categories over time that are different from the original definition and interpretation (Richards et al., 1999). Furthermore, behavior definition interpretations may change simultaneously and similarly among a pair of observers (O'Leary & Kent, 1973). Therefore, recreation of definition interpretation may not be represented in inter-observer agreement. Given that the observers were only trained initially, it is possible that one or more observers may have personally altered the interpretation of a comment category over the duration of the course; however, the observers physically consulted the category definitions (Appendix D) on each occasion that the observers rated student comments. Thus, the initial comment category

definitions were used as a point of reference for the duration of the course. This system should have minimized observer drift.

Mash and McElwee (1974) examined whether trained observers would more accurately rate verbal interactions using a four-point system than an eight-point system. The results of the study indicated that a higher level of inter-rater agreement was achieved when using the four category system than the eight category system. Kazdin (1977) cautions that the greater the number of categories, as well as the diversity of behaviors scored, can lower inter-observer agreement. In the current study, the observers used a 15-category observational system. Given findings from previous research, the complexity of the observational system may have reduced the inter-observer agreement. Kazdin (1977) recommends that observers participate in training sessions that incorporate all categories. The current study used a mock training session using a script and practicing in sections of the course not involved in the study. In both training sessions, all categories were included. Nonetheless, the complexity of the rating system may have contributed to poor inter-observer agreement for some categories.

Inter-observer agreement is more likely to be affected when the observer has an expectation of treatment outcomes and is provided feedback regarding the approval or disapproval of their ratings (O'Leary, Kent, & Kanowitz, 1975); however, expectancies alone do not alter inter-observer agreement (Kazdin, 1977). Thus, it is important to minimize the approval or disapproval of an observer's ratings. In the current study, no approval or disapproval feedback was provided to the primary or secondary observer. Only feedback regarding the accuracy of observers' ratings was provided, which is the method recommended by Kazdin.

Relationships between Course Variables and Participation Quality.

Students were classified into high- (90% productive) and low- (68% productive) quality groups based on the upper and lower quartile cut-offs of the average percent of productive comments made each discussion day. A discriminant-function analysis was used to determine which course variable or combination of course variables distinguished high- and low-quality groups the most. The course variables included the participation survey factors (Personal History and Confidence Regarding Participation, Expectation for Discussion in College Classes, and Personal Benefits of Participation), combined survey factors, total participation survey score, critical thinking scores, average unit exam scores, and average frequency of participation. Results of the discriminant-function analysis revealed that the combination of average daily frequency of participation and critical thinking scores significantly distinguished between students with high-quality participation and students with low-quality participation. Although no significant differences were found in critical thinking scores between the high- and low-quality groups, critical thinking in conjunction with the number of comments a student contributes on a daily basis classified 95% of students correctly into high- and low-quality groups. Specifically, students with higher critical thinking scores and higher daily participation frequency were more likely to be classified as a high-quality responder than a low-quality responder. McCleary et al. (2010) found critical thinking scores to be a significant predictor of participation frequency. It is likely that students with higher critical thinking skills who also comment frequently offer a higher percentage of productive comments than those with high critical thinking skills who do not contribute frequently to class discussion.

Further analysis of frequency of participation among qualitative levels revealed high-quality responders to participate significantly more than low-quality responders in all five units

in the sections combined. Proportion testing of the number of students in each qualitative group within each quantitative group indicated A-level students to be occasional or frequent responders (1 – 6 comments a day); B-level students tended to be dominant responders (7 or more comments a day); C-level students were equally likely to be occasional (1 – 3 comments per day), frequent (4 – 6 comments per day), or dominant responders (7 or more comments); and D-level students tended to be non-responders (0 comments per day). Although a student's qualitative level was independent of nonparticipation in that a student's qualitative level was based solely on the average percent of productivity on the days in which the student spoke, the results indicated that students disinclined to participate tended to contribute poor quality comments when they did comment in class. The relationship between quality of student participation and frequency of student participation may be a more curvilinear than linear. For example, if it were a linear relationship, we would expect A-level responders to be dominant responders; however, the results indicate that A-level responders actually tended to be occasional or frequent responders and B-level responders are predominantly dominant responders. Therefore, treatments aimed at balancing the quality and quantity of student participation may be most effective in targeting dominant and reticent responders. Scatterplot graphs are often used to assess the linearity of variable pairs (Tabackhnick & Fidell, 2007). Visual analysis of a scatterplot graph of the average number of comments contributed each discussion day and the average daily percent of productivity did not reveal a clear curvilinear trend. However, a larger sample size may produce a more pronounced visual trend.

Although the remaining course variables did not contribute to accurately distinguishing between low- and high-quality groups, some significant differences among the course variables occurred when each course predictor was included by itself. The Personal History and

Confidence Regarding Participation survey factor was significantly higher in the high-quality than the low-quality group. History/Confidence has been demonstrated to significantly predict frequency of student participation (Aspiranti, 2010; Krohn, 2010; Krohn et al., 2010). Also, low-quality responders scored significantly lower on the 23 combined factor items than high-quality responders. Students with high-quality participation indicated a combination of greater levels of personal history and confidence in their participation, higher expectations for discussion in college, and greater personal benefits of participation than low-quality responders. Furthermore, overall, high-quality responders significantly scored higher on unit exams than low-quality responders; however, a difference of this magnitude occurred only in two of five units. Inconsistency among participation quality groups in exam scores is likely a contributing factor for exam scores failing to accurately distinguish between high- and low- quality responders when exam scores were included with other course variables as predictors.

Secondary Research Findings

In addition to examining the reliability of the quality of participation rating system and relationships between course variables and participation quality, the researcher investigated the relationship of demographic variables (i.e., gender, academic classification, and grade point average) to participation variables, the effects of a random credit contingency on participation quantity, as well as the degree to which the participation contingency affected participation quality.

Demographic Variables and Participation Variables. There were no significant differences in frequency of participation between males and females, among academic classification levels, and grade point averages (GPA). Also, no differences in participation quality based on GPA or academic classification levels were found. However, females provided

significantly higher quality comments than males. Given that critical thinking scores may be related to participation quality, the difference in critical thinking scores between males and females was explored. A t-test yielded no significant difference in the critical thinking scores of males and females, $t(44) = 1.37$, $p = .178$. Therefore, females' tendency to contribute higher quality comments than males is likely due the proportion of the classes that were female. Males comprised only 25% of the entire class and were thus, in the minority. Findings regarding participation and gender are small in number and mixed. Females may have felt more confident in their responses, given that they were surrounded by similar peers.

Effects of a randomized credit contingency on frequency of participation. Students received credit for the number of comments they contributed on 2 randomly selected days in two of the five units. Students randomly and retrospectively chose the 2 days in each unit that would contribute toward their final grade at the end of each predetermined credit unit. The students received three points for their first comment and two points for an additional comment on each selected day. Therefore, a student who contributed more than two comments only received credit for the first two comments made. The inter-observer agreement between the primary observer and the secondary observer for the number of comments made ranged from 88 to 97%. The IOA between the primary observer and students ranged from 58 to 67% and 50 to 71% between the secondary observer and students' self-recording. The primary and secondary observer IOA was considerably higher than that between either observer and students. Examination of student over- and under-reporting demonstrated that students were equally likely to under-report their participation in credit and non-credit units and more likely to over-report their participation in credit than non-credit units. A minority of the over-reporting incidences resulted in undeserved full credit. Also, few students (3.4% of the students who under-reported during inter-rater days)

under-reported their participation below credit level participation (i.e., contributed one or two comments but did not record either comment). This pattern of under and over reporting are similar to the Aspiranti (2010), Krohn (2010), Krohn et al. (2009), and Krohn et al. (2010) studies. Given the somewhat inaccurate representation of student self-recorded participation, observer data were used in analyses involving frequency of participation.

Using participation levels similar to those identified in the Krohn et al. (2010) study, visual and proportion analyses revealed little evidence of a treatment effect. Additionally, there was no significant change in initially low, medium, or high responders within each quantitative level across course units. In general, students remained in the quantity category in which they were initially placed. Unlike previous studies (Aspiranti, 2010; Foster et al. 2009; Krohn, 2010; and Krohn et al.), which were conducted in large sections of this course, the current study consisted of small evening classes with about 25 students in each class. Thus, there were likely a greater percentage of nontraditional students and more opportunities for each student to respond than in large day sections. Balancing student participation in small sections composed partly of nontraditional students is more likely to be achieved with a higher credit level in the participation contingency. Given that students contributing one to three comments were designated as the occasional responder (i.e., the equivalent to the Krohn et al.'s credit-level participants), three comments would likely be a good starting place to examine the effects of a credit contingency on quantity of participation in small sections. Furthermore, most students made three to six comments in Unit 1; therefore, the majority of students responded at the credit-level in Unit 1.

Preserving participation quality while rewarding quantity. It is difficult to ascertain whether participation quality was preserved during units in which participation frequency was rewarded with credit due to the lack of a treatment effect on quantity of participation. Thus,

treatment effects on participation are likely muted in small classes and may be exaggerated in larger courses where class time is more restricted. Whatever the case, the absence of a treatment effect on quantity of participation pre-empts the possibility of determining the extent to which quantitative change indirectly produced qualitative changes.

Proportion testing that compared the proportion of productive comments made each discussion session in credit and non-credit units were all non-significant, with the exception of treatment pair comparisons that involved Unit 2. The average percent of productive comments made each discussion day significantly increased from Unit 2 to Unit 3 in both sections. Given that Unit 3 was a noncredit unit in Section A and a credit unit in Section B, I would expect higher-quality discussion in Unit 3 only in Section B, not in both sections, in order to establish a clear treatment effect on quality of discussion. One explanation for the decrease in the quality of the overall discussion in both sections in Unit 2 is that the percentage of students in lower qualitative levels was higher in both sections in Unit 2 than Unit 3 (See Figure 3), suggesting that the course material, which is typically more difficult in Unit 2 than Unit 3, probably accounted for the increase in quality of discussion sessions in Unit 3.

Visual inspection of the percent of A-level responders throughout the course revealed a slight treatment effect for Section B, although it was not statistically significant. The slight treatment effect for A-level students in Section B can probably be explained by lower exam scores in Section B than in Section A. Section B's (M = 37.06, SD = 5.39) overall average exam scores were significantly lower than Section A's (M = 40.57, SD = 4.67), $t(44) = 2.40$, $p = .021$. The quality of overall proportion of productive comments made in the entire class discussion was also significantly higher in combined credit units than in combined non-credit units in Section B but not in Section A (see Table 10). Even though individual student participation quality was

higher in treatment Units 3 and 5 in Section B than non-treatment Units 3 and 5 in Section A, Section B's exam scores were significantly lower in both treatment Units 3 and 5 than Section A. In addition, intra-section exam score comparisons revealed that Section B's exam scores were not significantly higher or lower in any given unit within Section B. Lower exam scores in Section B than A suggests that students in Section A may have been slightly better informed of course notions than students in Section B. This lack of course knowledge in Section B may have been due to an overall tendency for those students to spend an inadequate amount of time studying the course material. Therefore, providing students with a credit contingency may have encouraged students in Section B to be more prepared to discuss course materials than they otherwise would be, resulting in a slight increase in A-Level responders in credit units. Students in Section A may have been less affected by a credit contingency because they may have already had a tendency to be well informed and prepared for class discussion. There may be a link between increasing quality of discussion and exam scores; however, the current study failed to show an indirect linkage between quality of participation and exam scores.

Additionally, the percent of A-level responders decreased steadily from 81% in Day 1 to 30% in Day 4 in Unit 5 of Section B. This dramatic decrease is probably explained by the absence of frequently participating students on those days. The overall quality of discussion decreased from 95% on Day 1 to 78% on Day 4 in Unit 5 in Section B. Also, there were almost no dominant students (students who contributed seven or more comments each day) during these days. Given that students who talk more also make higher quality comments, it is possible that the dramatic decrease in Unit 5 of Section B is due to the absence of students who participated frequently. Additionally, there were also fewer dominating responders on the first day in each unit, yet percentage of A-level responders was not as affected as much as was the case on Day 4

of Unit 5. This may be due to a combination of opportunities to respond and course material difficulty. Day 1 in each unit was a video day, thus students had less time to contribute to discussion and very few students had the opportunity to contribute seven or more comments. Day 4 typically included discussion of articles, and had almost as many opportunities to respond as instructor note discussion days. However, the articles were usually less familiar than the information in the instructor notes. Also, the material discussed on Day 4 of Unit 5 was more controversial than material in earlier units. Furthermore, there may have also been a fatigue factor at the end of Unit 5.

Major Conclusions in the Current Study

In conclusion, the coding system used to measure the quality of participation produced inadequate inter-observer agreement in all but one of the 15 categories rated. Comprehension was the only category that yielded adequate inter-observer agreement in all five units. Poor inter-observer agreement was likely due to infrequent classification in many categories, especially the nonproductive categories that were only rated in combined sections by any one observer 1 to 12 times per inter-rater day. In combined sections, six of the seven non-productive categories were rated by one and/or the other observer 1 to 9 times per inter-rater day. The poor comprehension category was rated 6 to 12 times by an observer across inter-rater days in combined sections. Thus, all but one non-productive category was observed on less than 10 occurrences across sections. Investigation of the inter-observer agreement of the total number of productive comments produced more consistent and acceptable levels of agreement (i.e., 84 to 95% across all units in combined sections). Inter-observer agreement on the total number of nonproductive comments contributed by each student was lower than for productive comments; thus, mitigating against adequate inter-observer agreement for non-productive comments.

Therefore, the total number of productive comments contributed by each student was used in the calculation of percent of productivity (i.e., total number of productive comments divided by total number of comments contributed). Given the strong reliability of inter-rater agreement for the percentage of productive comments in the total number of comments contributed by each student, percent of daily productivity was used as a major indicator of student quality of participation.

Additionally, average daily frequency of participation and critical thinking scores were significant predictors of high- and low-quality responders such that students who both contributed more comments to class participation and scored higher on a measure of critical thinking were more likely to be a high quality responder than students making fewer comments and scoring lower on critical thinking. Furthermore, inspection of quantitative levels within each qualitative level across units revealed A-level responders to be occasional or frequent responders, B-level responders to be dominant responders, C-level responders to be equally likely to be occasional, frequent, or dominant responders, and D-level responders to be non-responders. Thus, high-quality responders tended to be moderate in the number of comments they contributed to class discussion. Students who contributed the most frequently to class discussion tended to be composed of B- or C-level responders.

Although about half of students under-reported the number of comments they contributed each day, students did not under-report their participation more or less in credit than non-credit units. However, students did over-report their participation more during credit-units than non-credit units. The majority of students (six out of eight students) who over-reported their participation to receive partial credit only did so on 1 out of 8 credit days. Regardless, inter-observer agreement of quantity of student participation ranged from 88 to 97% across the five

units. Visual inspection and proportion analysis of quantity of student participation across credit and non-credit units yielded little to no evidence of a treatment effect, thus the researcher was unable to determine any effect that the credit contingency would have on quality of participation.

Limitations and Directions for Future Research

A major limitation of the current study was failure to find a treatment effect for the frequency of student participation. Unlike previous studies (Aspiranti, 2010; Foster et al., 2009; Krohn, 2010; Krohn et. al., 2010), the credit contingency in the current study, which awarded students three points for their first comment and two points for an additional comment on 2 days randomly chosen at the end of preannounced credit units, did not contribute to changes in the percent of non-responders, credit-level responders, or dominant responders. Therefore, the researcher was unable to track how changes in quantity of participation affected quality of participation. Failure to balance participation in credit units was likely due to more opportunities for students to respond than those provided in previous studies (Aspiranti; Foster et al.; Krohn; Krohn et al.). The sections included in the current study were 75 minutes in duration and consisted of about 25 students each. In comparison to the sections included in previous studies, which were 50 minutes in duration and consisted of 55 students each, the opportunity for student contribution to class discussion in the current study was far greater. Also, students in smaller college courses, as well as non-traditional students (i.e., over age 25), tend to participate more in class discussion (Fassinger, 1995; Howard & Henney, 1998; Weaver & Qi, 2005). It is likely that the sections in the current study, which occurred after 5 P.M. on 2 days a week, were composed of more non-traditional students than the sections used in previous studies, which occurred between 10:10 A.M. and 1:20 P.M. three days a week.

Providing students credit for only two comments was not a strong treatment to balance student participation. The percentage of students who made two or more comments on the first 4 days in Unit 1 of the current study ranged from 62 to 82%. Thus, the majority of students contributed to class discussion at treatment level before the credit contingency was put in place. In order to increase the likelihood of producing a treatment effect on the amount of student participation in college courses that meet in the evening with a similar number of students, the credit contingency should require students to contribute at least three to four comments a class period for credit. Previous studies (Aspiranti; Foster et al.; Krohn; Krohn et al.) have demonstrated that student participation is more balanced in units in which 5 to 6% of course credit is offered for participation than units in which no credit is given for participation. Therefore, instructors may need to make each comment worth less than 2 or 3 points a comment, as well as offer at least 5% of course credit for participation in order to encourage participation above baseline levels in small sections.

The current study is also limited in some aspects of external validity. The findings of the study may be generalized to college courses composed of similar students (i.e., enrolled in or preparing to enroll in a teacher-education program). Furthermore, the findings are only generalizable to college courses with similar course logistics aimed at preparing students for high quality class discussions (i.e., reading and answering questions over the days' discussion material prior to class discussion, possessing all course materials in advance). Also, the instructor of both sections in the current study was trained to ask higher-ordered questions, limit lecturing, and maximize discussion time. Given that many college course instructors employ a traditional lecture style, it is important to highlight the potential importance of methods used to maximize and increase the quality of class discussion in the current study. Future research needs

to be conducted to ascertain which instructor and/or course logistic components are most valuable in increasing quality of class discussion and ultimately have the greatest impact on learning.

The current study also had a number of procedural limitations: lack of monitoring of instructor behaviors, researcher bias, and inadequate inter-observer agreement of quality of participation categories. Previous studies (Aspiranti, 2010; Krohn, 2010; Krohn et al.; 2010) that have investigated student participation have also monitored teacher behaviors (i.e., the number of lower ordered and higher order questions asked and the amount of positive and negative feedback given). Student participation may be dependent on or related to instructor behaviors (Auster & MacRone, 1994); thus, it is important to monitor the instructor behaviors in each section to ensure that differences in student participation are not due to or are a reaction to instructor behaviors. Krohn (2010) found that instructors tended to ask more questions in sections in which students were more reticent. Due to a lack of resources (i.e., observers to monitor teacher behaviors), the current study was unable to ensure that qualitative changes were not associated with instructor question type or feedback. Although the current study did not find significant or clear participation patterns, it is important for future researchers to monitor instructor behaviors, especially when the credit contingency is strong enough to produce significant changes in student participation levels.

A potential weakness and subsequent threat to internal validity in the current study is that the primary investigator of the study was also the primary observer. The primary investigator was also aware of the specifics of the credit contingencies. Typically, when possible researcher bias, also known as experimenter bias, is present, the researcher may be more likely to produce findings that are favorable to the researcher's hypothesis. The current study did not produce a

clear and significant treatment effect; therefore, it is unlikely that researcher bias was present. Nonetheless, future research investigating the quality of student participation should attempt to eliminate researcher bias by excluding the primary investigator as an observer and ensuring that observers are blind to credit contingencies.

Given the low observation of some nonproductive and productive categories, it may be beneficial to collapse those particular categories (i.e., Skeptical, Locate, Poor Testimony, Reading Answers, and Disagreeing Negatively) into the others in order to simplify the rating system. Also, comments categorized under the Comprehension category included extremely high-quality comments (i.e., comments that synthesized course concepts correctly), as well as high-quality comments that included plausible student explanations for a research finding. Thus, the comprehension category included student comments that fell under all but one of Bloom's (1956) taxonomy of learning categories. Therefore, the comprehension category may have been too broad. It may be beneficial to separate comments that fell under the Comprehension category in the current study into two categories. This would allow for a distinction within the productive category that differentiates good quality comments from superior comments, which would ultimately create more variability in quality participation. A similar distinction in productive comments was also made in Mainkar's (2007) system of evaluating student comments. Mainkar delineated three types of student comments: incorrect, straightforward, and insightful. Reducing the rating system to three or four categories would simplify the coding system used in the current study and might allow for greater accuracy of inter-observer agreement (Kazdin, 1977).

In addition, it would be beneficial to offer multiple training sessions to protect against observer drift. Furthermore, the accuracy of inter-observer agreement might be strengthened if

the observers were not aware of which days counted for inter-observer agreement. Finally, it is ideal to rate quality of comments from a typed transcript in order to ascertain where a comment begins and ends. Disagreement between observers sometimes relates to where a comment begins and ends. Although inter-observer agreement of the number of comments that each student contributes is considered high, I could not be sure that observers were recording the exact same comment. Thus, the use of a typed transcript based on an auditory or visual recording of the class would allow for both inter-rater agreement of number of comments contributed and quality of comments. Once observers are able to rate identical behaviors, the inter-observer agreement may be assessed using the kappa coefficient. The kappa coefficient not only accounts for the probability of agreement due to chance but it also is not greatly affected by the frequency of occurrence (Byrington et. al., 2002).

List of References

Aspiranti, K. B. (2010). *Random credit of participation in college classes*. Unpublished doctoral dissertation, University of Tennessee, Knoxville.

Auster, C. J., & MacRone, M. (1994). The classroom as a negotiated social setting: An empirical study of the effects of faculty members' behavior on students' participation. *Teaching Sociology, 22*, 289-300.

Barnes, C. P. (1983). Questioning in college classrooms. In C. L. Ellner & C. P. Barnes (Eds.), *Studies of college teaching* (pp. 61-82). Lexington, MA: D. C. Heath.

Bean, J. C., & Peterson, D. (1998). Grading classroom participation. *New Directions for Teaching and Learning, 74,* 33-40. doi: 10.1002/tl.7403

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives, handbook I: Cognitive domain*. New York: David McKay.

Boniecki, K. A., & Moore, S. (2003). Breaking the silence: Using a token economy to reinforce classroom participation. *Teaching of Psychology, 30*, 224-227. doi: 10.1207/s15328023t0p3003_05

Bruss, K. (2009). Improving classroom discussion: A rhetorical approach. *The Journal of General Education, 58*, 28-46. doi: 10.1353/jge.0.0034

Bryington, A. A., Plamer, D. J., & Watkins, M. W. (2002). The estimation of interobserver agreement in behavioral assessment. *The Behavior Analyst Today*, *3*, 323-326.

Burchfield, C. M., & Sappington, J. (1999). Participation in classroom discussion. Teaching of Psychology, 26, 277-294. doi: 10.1207/s15328023t0p3003_05

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Erlbaum.

Fassinger, P. A. (1995). Professors' and students' perceptions of why students participate in class. *Teaching Sociology, 24,* 25-33.

Ferguson, G. A., & Takane, Y. (1989) Statistical Analysis in Psychology and Education. New York: McGraw-Hill

Foster, L. N., Krohn, K. R., McCleary, D. F., Aspiranti, K. B., Nalls, M. L., Quillivan, C. C., et al. (2009). Increasing low-responding students' participation in class discussion. *Journal of Behavioral Education, 18,* 173-188. doi: 10.1007/s10864-009-9083-8

Garside, C. (1996). Look who's talking: A comparison of lecture and group discussion teaching strategies in developing critical thinking skills. *Communication Education, 45,* 212-227. doi: 10.1080/03634529609379050

Garson, G. D. (2008). *Scales and Standard Measures* from Statnotes: Topics in Multivariate Analysis. Retrieved December 28, 2010, from http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm.

Gilson, C. (1994). Of dinosaurs and sacred crows: The grading of classroom participation. *Journal of Management Education, 18*, 227-236.

Gopinath, C. (1999). Alternatives to instructor assessment of class participation. *Journal of Management Education, 18,* 227-236. doi: 10.1080/08832329909598983

Hodge, G. K., & Nelson, N. H. (1991). Demonstrating differential reinforcement by shaping classroom participation. *Teaching of Psychology, 18*, 239-241. doi: 10.1207/s15328023top1804_13

House, A. E., Farber, J. W., & Nier. L. L. (1983). Differences in computational accuracy and speed of calculation between three measures of interobserver agreement. *Child Study Journal*, *13*, 195-201.

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

Howard, J. R., James, G. H., III, & Taylor, D. R. (2002). The consolidation of responsibility in the mixed-age college classroom. *Teaching Sociology, 30,* 214-234.

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

Kazdin, A. E. (1977). Artifact, bias, and complexity of assessment: The ABC's of reliability. *Journal of Applied Behavior Analysis*, *10*, 141-150. doi: 10.1901/jaba.1977.10-141

Kratochwill, T. R., & Weltzel, R. J. (1977). Observer agreement, credibility, and judgement: Some considerations in presenting observer agreement data. *Journal of Applied Behavior Analysis*, *10*, 133-139. doi: 10.1901/jaba.1977.10-133

Krohn, K. R. (2010). *The effect of self-recording and contingent credit on the quantity and relevance of student participation in class discussion in large college classes*. Unpublished doctorial dissertation, University of Tennessee, Knoxville.

Krohn, K. R., Aspiranti, K. B., Foster, L. N., McCleary, D. F., Taylor, C. M., Nalls, M. L., Quillivan, C. C., & Williams, R. L. (2010). Effects of self-recording and contingent credit on balancing participation across students. *Journal of Behavioral Education, 19*, 134-155. doi: 10.1007/s10864-010-9105-6

Love, K. B. (1981). Comparison of peer assessment methods: Reliability, validity, friendship, bias, and user reaction. *Journal of Applied Psychology, 66*, 451-457. doi:10.1037/0021-9010.66.4.451

Mainkar, A. V. (2007). A student-empowered system for measuring and weighing participation

in class discussion. *Journal of Management Education, 32*, 23-37. doi: 10.1177/1052562906286912

Mash, E. J., & McElwee, J. (1974). Situational effects on observer accuracy: Behavioral predictability, prior experience, and complexity of coding categories. *Child Development*, *45*, 367-377.

Melvin, K. B. (1988). Rating class participation: The prof/peer method. *Teaching of Psychology, 3*, 137-139. doi: 10.1207/s15328023top1503_7

McCleary, D. F., Foster, L.N., & Williams, R.L. (2010). *Relationship of class participation to critical thinking and test performance.* Manuscript submitted for publication.

O'Leary, K. D., & Kent, R. N. (1973). Behavior modification of social action: Research tactics and problems. In L. A. Hamerlynck, P. O. Davidson, I. E. Acker (Eds.), *Critical Issues in Research and Practice,* (pp. 69 – 96). Champaign, IL: Research Press.

O'Leary, K. D., Kent, R. N., & Kanowitz, J. (1975). Shaping data collection congruent with experimental hypotheses. *Journal of Applied Behavior Analysis*, *8*, 43-51. doi: 10.1901/jaba.1975.8-43

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

Richards, S. B., Taylor, R. L., Ramasamy, R., & Richards, R. Y. (1999). Single subject research: Applications in educational and clinical settings. London: Wadsworth.

Reid, J. B. (1970). Reliability assessment of observation data: A possible methodological problem. *Child Development*, *41*, 1143-1150.

Ryan, G. J., Marshall, L. L., Porter, K. P., & Jia, H. (2007). Peer, professor and self-evaluation of

class participation. *Active Learning in Higher Education, 8*, 49-61. doi: 10.1177/1469787407074049

Smith, D. G. (1977). College classroom interactions and critical thinking. *Journal of Educational Psychology, 69,* 180-190. doi:10.1037/0022-0663.69.2.180

Sommer, R., & Sommer, B. A. (2007). Credit for comments, comments for credit. *Teaching of Psychology, 34*, 104-106. doi: 10.1207/s15328023top2802_06

Stitt, J. K., Simonds, C. J., & Hunt, S. K. (2003). Evaluation fidelity: An examination of criterion-based assessment and rater training in the speech communication classroom. *Communication Studies, 54*, 341-351.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics, Fifth Edition*. Boston: Pearson Education, Inc.

Wallace, M. A., & Williams, R. L.(2003). Multiple choice exams: Explanations for student choices. *Teaching of Psychology, 30*, 136-138.

Watson, G. B., & Glaser, E. M. (1994). *WGCTA: Watson-Glaser Critical Thinking Appraisal.* San Antonio: The Psychological Corporation.

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

Williams, R. L., Oliver, R., Allin, J. L., Winn, B., & Booher, C. S. (2003a). Psychological critical thinking as a course predictor and outcome variable. *Teaching of Psychology, 30,* 220-223. doi: 10.1207/S15328023TOP3003_04

Williams, R. L., Oliver, R., Allin, J. L., Winn, B., & Booher, C. S. (2003b). Knowledge and critical thinking as course predictors and outcomes. *Inquiry: Critical Thinking across the Disciplines, 22,* 57-63.

Williams, R. L., Oliver, R., & Stockdale, S. (2004). Psychological versus generic critical thinking as predictors and outcome measures in a large undergraduate human development course. *Journal of General Education, 53*, 37-58. doi: 10.1353/jge.2004.0022

Williams, R. L., & Stockdale, S. L. (2003). High-performing students with low critical thinking skills. *Journal of General Education, 52,* 200-226. doi: 10.1353/jge.2004.0007

Williams, R. L., & Worth, S. L. (2002). Thinking skills and work habits: Contributors to course performance. *The Journal of General Education, 51,* 200-227. doi: 10.1353/jge.2003.0007

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

Appendices

Appendix A: Tables

Table 1

Treatment Conditions Assigned Across the Two Sections

Section	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5
A (5:05)	Baseline	RC^{2x}	NC	RC^{2x}	NC
B (6:30)	Baseline	NC	RC^{2x}	NC	RC^{2x}

Note. RC^{2x} = treatment phase in which 2 days were randomly drawn for participation credit at the end of the unit. NC = non-treatment phase in which students did not receive credit for participation.

Table 2

Mean and Standard Deviations of Daily Percent of Productive Comments Based on Student Gender and Academic Classification

	Mean Daily Percent of Productive Comments		
	Overall	CR^{2x}	NC
Gender			
Male (*n* = 15)	62.4 (28.0)	67.9 (34.9)	54.3 (33.2)
Female (*n* = 30)	82.4 (14.5)	88.6 (12.6)	74.2 (24.0)
Academic Classification			
Freshman (*n* = 2)	81.6 (7.2)	75.0 (1.8)	84.5 (17.3)
Sophomore (*n* = 18)	71.8 (24.2)	80.4 (24.9)	58.5 (33.3)
Junior (*n* = 14)	81.2 (16.0)	86.6 (15.5)	74.4 (22.9)
Senior (*n* = 8)	67.1 (27.9)	73.5 (39.4)	62.3 (27.3)
Graduate (*n*=3)	92.8 (7.5)	93.6 (7.1)	92.9 (6.5)

Note. Numbers in parentheses represent standard deviations. Overall = the daily frequency of participation mean computed with all five units. CR^{2x} = the daily participation mean in credit units. NC = the daily participation mean in non-credit units.

Table 3

Mean and Standard Deviations of Number of Comments Students Contributed Based on Student Gender and Academic Classification

	Daily Participation Mean		
	Overall	CR^{2x}	NC
Gender			
Male (*n* = 15)	3.21 (3.30)	3.25 (3.04)	3.10 (3.84)
Female (*n* = 30)	4.24 (2.76)	4.46 (2.95)	4.19 (3.23)
Academic Classification			
Freshman (*n* = 2)	4.83 (0.15)	4.10 (0.38)	5.92 (0.35)
Sophomore (*n* = 18)	3.33 (2.55)	3.45 (2.34)	3.03 (3.12)
Junior (*n* = 14)	3.72 (2.56)	3.88 (2.23)	3.58 (2.93)
Senior (*n* = 8)	3.80 (3.60)	4.30 (4.43)	3.80 (3.84)
Graduate (*n* = 3)	7.73 (4.75)	7.86 (4.98)	8.44 (5.26)

Note. Numbers in parentheses represent standard deviations. Overall = the daily frequency of participation mean computed with all five units. CR^{2x} = the daily participation mean in credit units. NC = the daily participation mean in non-credit units.

Table 4

Percent of Agreement between Primary and Secondary Raters for Productive Categories

	Unit					
	1	2	3	4	5	Average
Section A	b(*n*)	c (*n*)	nc(*n*)	c(*n*)	nc(*n*)	
Comprehension	83 (22)	69 (20)	81 (19)	80 (15)	67 (11)	76 (17)
Answer	38 (8)	0 (14)	22 (3)	NR	25 (4)	21 (7)
Factual	57 (7)	59 (11)	52 (8)	63 (9)	49 (7)	56 (8)
Skeptical	0 (2)	0 (1)	NR	0 (1)	0 (2)	0 (2)
Locate	0 (2)	100 (1)	NR	0 (1)	NR	33 (1)
Clarify	38 (4)	50 (4)	29 (7)	50 (2)	83 (3)	50 (4)
Testimony	71 (13)	39 (6)	85 (10)	80 (4)	67 (4)	68 (7)
Other Productive	NR	50 (2)	NR	0 (1)	NR	25 (2)
Total Productive	98 (23)	91 (20)	91 (19)	93 (15)	89 (12)	92 (18)
Section B	b(*n*)	nc(*n*)	c(*n*)	nc(*n*)	c(*n*)	
Comprehension	81 (17)	73 (15)	89 (17)	91 (16)	81 (17)	83 (16)
Answer	21 (7)	25 (8)	NR	NR	0 (3)	15 (6)
Factual	63 (9)	38 (8)	90 (8)	53 (12)	50 (9)	59 (9)
Skeptical	NR	NR	NR	NR	NR	NR
Locate	NR	NR	NR	NR	NR	NR
Clarify	0 (1)	0 (1)	NR	0 (2)	100 (2)	25 (2)
Testimony	66 (13)	40 (5)	60 (9)	50 (6)	100 (1)	63 (9)
Other Productive	NR	NR	NR	NR	NR	NR
Total Productive	91 (19)	85 (16)	90 (17)	89 (16)	81 (18)	87 (17)
Combined Sections						
Comprehension	82 (39)	71 (35)	85 (36)	86 (31)	75 (28)	80 (34)
Answer	30 (15)	9 (22)	22 (3)	NR	14 (7)	19 (12)
Factual	60 (16)	50 (19)	71 (16)	57 (21)	49 (16)	57 (18)
Skeptical	0 (2)	0 (1)	NR	0 (1)	0 (2)	0 (2)
Locate	0 (2)	100 (1)	NR	0 (1)	NR	33 (1)
Clarify	30 (5)	40 (5)	29 (7)	25 (4)	90 (5)	43 (5)
Testimony	68 (26)	39 (11)	73 (19)	62 (10)	73 (5)	63 (14)
Other Productive	NR	50 (2)	NR	0.00 (1)	NR	25 (2)

Table 4 continued

Total Productive	95 (42)	89 (36)	91 (36)	91 (31)	84 (30)	90 (35)

Note. NR = no instances of rating under a category on a particular day. *n* = highest number of occurrences for any given observer

Table 5

Correlations Representing Inter-Rater Agreement for Productive and Nonproductive Comments

	Unit				
	1	2	3	4	5
Section A	(n = 24)[b]	(n = 23)[c]	(n = 22)[nc]	(n = 18)[c]	(n = 19)[nc]
Productive	.99	.96	.99	.99	.96
Nonproductive	.74	.74	.82	.58*	.77
Section B	(n = 20)[b]	(n = 17)[nc]	(n = 18)[c]	(n = .17)[nc]	(n = 20)[c]
Productive	.98	.98	.97	.96	.96
Nonproductive	.53*	.93	.74	.96	.64*
Combined	(n = 44)	(n = 40)	(n = 40)	(n = 35)	(n = 39)
Productive	.98	.98	.98	.97	.96
Nonproductive	.62	.82	.78	.86	.73

[b] = baseline unit. [c] = credit unit. [nc] = noncredit unit.

*$p < .05$; all other correlations significant at the $p < .001$ level.

Table 6

Percent of Agreement between Primary and Secondary Raters for Non-Productive Comments

	Unit					
	1	2	3	4	5	Average
Section A	b(*n*)	c(*n*)	nc(*n*)	c(*n*)	nc(*n*)	
Poor Comprehension	13 (4)	64 (7)	42 (6)	13 (4)	44 (3)	29 (5)
Poor Factual	100 (2)	25 (2)	50 (3)	100 (1)	30 (4)	51 (2)
Poor Testimony	NR	NR	0 (1)	NR	NR	0 (1)
Poor Joke	100 (1)	0 (1)	NR	NR	100 (2)	50 (1)
Reading Answers	0 (1)	NR	0 (1)	100 (1)	0 (1)	20 (1)
Disagree Negatively	NR	100 (1)	NR	NR	0 (1)	33 (1)
Other	NR	NR	0 (1)	NR	0 (1)	0 (1)
Total Nonproductive	64 (7)	52 (9)	57 (10)	42 (6)	65 (7)	41 (8)
Section B	b(*n*)	nc(*n*)	c(*n*)	nc(*n*)	c(*n*)	
Poor Comprehension	17 (6)	50 (5)	40 (5)	100 (2)	40 (5)	41 (5)
Poor Factual	NR	100 (3)	20 (5)	100 (1)	50 (5)	54 (3)
Poor Testimony	NR	NR	NR	NR	NR	NR
Poor Joke	100 (1)	100 (2)	17 (3)	39 (3)	100 (1)	59 (2)
Reading Answers	NR	NR	NR	NR	NR	NR
Disagree Negatively	NR	NR	NR	NR	NR	NR
Other	NR	NR	NR	NR	NR	NR
Total Nonproductive	25 (6)	75 (7)	63 (8)	68 (5)	81 (8)	52 (7)
Combined						
Poor Comprehension	15 (10)	58 (12)	41 (11)	42 (6)	41 (8)	33 (9)
Poor Factual	100 (2)	70 (5)	31 (8)	100 (2)	39 (9)	57 (5)
Poor Testimony	NR	NR (0)	0 (1)	NR	NR	0 (1)
Poor Joke	100 (2)	67 (3)	30 (5)	39 (3)	100 (3)	56 (3)
Reading Answers	0 (1)	NR	100 (1)	100 (1)	0 (1)	40 (1)
Disagree Negatively	NR	100 (1)	NR	NR	0 (1)	33 (1)
Other	NR	NR	0 (1)	NR	0 (1)	0 (1)
Total Nonproductive	46 (13)	62 (16)	59 (18)	54 (11)	74 (31)	49 (18)

Table 6 continued

Note. NR = no instances of rating under a category on a particular day. *n* = highest number of occurrences for any given observer

Table 7

Average Inter-Observer Percent Agreement between Student and Primary Observer Records of Number of Comments on Baseline (b), Credit (c), and Non-Credit (nc) Units

	Units				
	1	2	3	4	5
Section A	[b]	[cr]	[nc]	[cr]	[nc]
Day 1	73	79	69	70	62
Day 2	63	70	35	67	48
Day 3	72	64	55	69	48
Day 4	74	77	54	72	48
Section B	[b]	[nc]	[cr]	[nc]	[cr]
Day 1	71	53	92	51	66
Day 2	62	66	64	66	65
Day 3	61	51	57	54	62
Day 4	62	55	68	45	69
All Sections					
Day 1	72	70	81	63	64
Day 2	63	68	51	67	58
Day 3	67	58	56	61	56
Day 4	68	67	60	63	58

[b] = baseline days. [cr] = credit days. [nc] = non-credit days.

Table 8

Average Inter-Observer Percent Agreement (IOA) between Student and Observer Records of Number of Comments on Inter-rater Check in Baseline (b), Credit (c), and Non-Credit (nc) Units

	Units				
Pairs within sections	1	2	3	4	5
Section A	(n = 23)[b]	(n = 20)[cr]	(n = 19)[nc]	(n = 15)[cr]	(n = 13)[nc]
Students and observer 1	72	64	55	69	48
Students and observer 2	73	67	54	69	48
Observers 1 and 2	98	92	98	91	96
Section B	(n = 19)[b]	(n = 16)[nc]	(n = 17)[cr]	(n = 16)[nc]	(n = 18)[cr]
Students and observer 1	61	51	57	54	62
Students and observer 2	86	51	57	54	79
Observers 1 and 2	96	88	92	88	84
All sections	(n = 42)	(n = 36)	(n = 36)	(n = 31)	(n = 31)
Student and observer 1	67	58	56	61	56
Student and observer 2	71	50	56	57	55
Observers 1 and 2	97	91	95	89	88

[b] = baseline days (no credit or self-recording). [c] = credit units. [nc] = non-credit units. n = highest number of occurrence for any given observer.

Table 9

Differences in Proportions of Students at Different Qualitative Levels between Treatment Conditions Overall and between Pairs of Adjacent Treatment Units in Each Section

	Comparisons		
	Overall	Treatment Pair 1	Treatment Pair 2
Section A	(C v. NC)	(C v. NC)	(C v. NC)
A level	14/18 > 12/17; p = .42	13/20 < 13/17; p = .22	14/18 < 11/14; p = .49
B level	0/19 = 0/17; NP	0/20 = 0/17; NP	0/18 = 0/14; NP
C level	4/19 > 2/17; p = .23	4/20 > 2/17; p = .25	3/18 > 1/14; p = .21
D level	2/19 < 2/17; p = .45	3/20 > 1/17; p = .19	0/18 < 2/14; p = .05
Section B	(C v. NC)	(NC v. C)	(NC v. C)
A level	12/17 > 10/15; p = .41	10/14 < 14/17; p = .23	7/12 < 11/16; p = .28
B level	0/17 < 1/15; p = .14	1/14 > 1/17; p = .44	1/12 > 1/16; p = .42
C level	2/17 < 2/15; p = .45	3/14 > 1/17; p = .10	2/12 > 2/16; p = .38
D level	2/17 < 2/15; p = .45	1/14 > 1/17; p = .44	2/12 > 2/16; p = .38

Note. NP = no proportion was calculated due to equal proportions. C = Credit. NC = Noncredit.

Table 10

Differences in Proportions of Productive Comments Made Each Discussion Session between Treatment Conditions Overall and between Pairs of Adjacent Treatment Units in Each Section

	Comparisons
Section A	
Noncredit vs. Credit Units	856/913 > 663/718; $p = .13$
Treatment pair 1 (C v. NC)	326/361 < 316/332; $p = .007$
Treatment pair 2 (NC v. C)	337/357 > 241/259; $p = .25$
Section B	
Noncredit vs. Credit Units	812/898 < 577/616; $p = .01$
Treatment Pair 1 (NC v. C)	263/285 < 332/347; $p = .04$
Treatment Pair 2 (C v. NC)	280/316 < 245/269; $p = .16$

Table 11

Differences in Proportions of Students at Different Quantitative Levels between Treatment Conditions Overall and between Pairs of Adjacent Treatment Units in Each Section

	Comparisons		
	Overall	Treatment Pair 1	Treatment Pair 2
Section A	(C v. NC)	(C v. NC)	(C v. NC)
Non-responder	3/22 < 5/21; p = .20	3/23 < 5/22; p = .20	3/21 < 7/20; p = .06
1 – 3 Comments	10/22 > 8/21; p = .31	11/23 > 8/22; p = .22	9/21 > 7/20: p = .30
4 – 6 Comments	4/22 < 4/21; p = .47	5/23 > 4/22; p = .38	4/21 < 4/20; p = .47
Dominant	5/22 > 4/21; p = .38	4/23 < 5/22; p = .33	5/21 > 3/20; p = .24
Section B	(C v. NC)	(NC v. C)	(NC v. C)
Non-responder	2/18 < 4/18; p = .19	4/18 > 2/18; p = .19	5/17 > 3/19; p = .16
1 – 3 Comments	8/18 > 7/18; p = .33	7/18 < 8/18: p = .97	5/17 < 10/19; p = .08
4 – 6 Comments	4/18 > 3/18; p = .34	3/18 < 4/18; p = .34	3/17 < 4/19; p = .40
Dominant	4/18 = 4/18; p = NP	4/18 = 4/18; NP	5/17 > 3/19; p = .16

Note. NP = no proportion was calculated due to equal proportions; C = Credit; NC = Noncredit.

Table 12

Number of Low-Responding Students in Unit 1 Who Fell into Low, Medium, and High Categories in Subsequent Units

	Units				
	1	2	3	4	5
Section A	5 [b]	5 [c]	5 [nc]	5 [c]	5 [nc]
Low	5	5	4	4	5
Medium	0	0	1	1	0
High	0	0	0	0	0
Section B	7 [b]	7 [nc]	7 [c]	7 [nc]	7 [c]
Low	7	4	3	5	4
Medium	0	3	4	1	2
High	0	0	0	1	1

Note. Low = 2 comments a day or less. Medium = 3 to 5 comments a day. High = 6 comments a day or more.

[b] = baseline units. [c] = credit units. [nc] = non-credit units.

Table 13

Number of Medium-Responding Students in Unit 1 Who Fell into Low, Medium, and High Categories in Subsequent Units

	Units				
	1	2	3	4	5
Section A	13 [b]	13 [c]	13 [nc]	13 [c]	13 [nc]
Low	0	1	5	3	8
Medium	13	10	4	7	3
High	0	2	4	3	2
Section B	7 [b]	7 [nc]	7 [c]	7 [nc]	7 [c]
Low	0	2	0	1	2
Medium	7	4	2	4	5
High	0	1	5	2	0

Note. Low = 2 or less comments a day. Medium = 3 to 5 comments a day. High = 6 or more comments a day.

[b] = baseline units. [c] = credit units. [nc] = non-credit units.

Table 14

Number of High-Responding Students in Unit 1 Who Fell into Low, Medium, and High Categories in Subsequent Units

	Units				
	1	2	3	4	5
Section A	6 [b]	6 [c]	6 [nc]	6 [c]	6 [nc]
Low	0	0	0	0	0
Medium	0	2	1	1	3
High	6	4	5	5	3
Section B	7 [b]	7 [nc]	7 [c]	7 [nc]	7 [c]
Low	0	1	0	0	0
Medium	0	1	3	2	2
High	7	5	4	5	5

Note. Low = 2 or less comments a day. Medium = 3 to 5 comments a day. High = 6 or more comments a day.

[b] = baseline units. [c] = credit units. [nc] = non-credit units.

Table 15

Mean Scores and Standard Deviations on the Participation Survey for Low- and High-Quality Participating Students

	Participation Groups	
	Low	High
Total Survey[a]	171.56 (16.72)	184.70 (13.34)
Combined Factors[b]	74.82 (12.46)*	85.70 (10.83)
Factor 1: History/Confidence[c]	24.91 (5.67)*	30.50 (4.33)
Factor 2: Expectation[c]	25.91 (4.81)	28.50 (4.67)
Factor 3: Personal Benefits[d]	24.00 (3.90)	26.70 (3.68)
Discarded Survey Items[e]	96.73 (7.64)	99.00 (5.85)

[a]Possible score range was 50 - 250. [b]Possible score range was 23 – 115. [c]Possible score range was 8 – 40. [d]Possible score range was 7 – 35. [e]Possible score range was 27 – 135.

*Significant difference between Low and High, $p < .05$

Table 16

Mean Scores and Standard Deviations on Unit Exams for the Unit's Low- and High-Quality Participating Students

	Participation Groups	
	Low	High
Overall Exam Scores[a]	37.52 (5.01)*	42.08 (3.92)
Unit 1[b]	36.11 (4.96)*	40.80 (4.48)
Unit 2[c]	35.29 (6.13)	41.00 (6.98)
Unit 3[d]	42.40 (3.47)	40.21 (5.59)
Unit 4[e]	34.55 (9.00)*	39.91 (5.41)
Unit 5[f]	40.73 (4.65)	40.38 (5.56)

[a] Low (n = 11) High (n = 10); [b] Low (n = 9) High (n = 20); [c] Low (n = 7) High (n = 20); [d] Low (n = 10) High (n = 24); [e] Low (n = 11) High (n = 22); [f] Low (n = 11) High (n = 13).

* = significant difference between Low and High, $p < .05$.

Table 17

Mean Scores and Standard Deviations of Average Participation Frequency per day for each Unit's Low- and High-Quality Participating Students

	Participation Groups	
	Low	High
Overall Course[a]	1.07 (.78)*	5.44 (3.15)
Unit 1[b]	2.64 (1.93)*	5.38 (3.09)
Unit 2[c]	1.30 (1.07)*	5.11 (3.04)
Unit 3[d]	1.30 (1.24)*	6.25 (3.91)
Unit 4[e]	2.28 (3.11)*	6.17 (4.01)
Unit 5[f]	1.92 (1.90)*	5.46 (4.55)

[a] Low (n = 11) High (n = 10); [b] Low (n = 9) High (n = 20); [c] Low (n = 7) High (n = 20); [d] Low (n = 10) High (n = 24); [e] Low (n = 11) High (n = 22); [f] Low (n = 11) High (n = 13).

* = significant difference between Low and High, $p < .05$.

Table 18

Discriminant-Function Analysis Results of Course Variables and Qualitative Groups

Predictor Variable	Correlations of Predictor Variables with Discriminant Function	Canonical DF Coefficients	
		Standardized	Unstandardized
Daily Course Participation[a]	.76	1.14	.51
Average Exam Score	.33		
Personal Benefits	.24		
Critical Thinking[a]	.17	.75	.03
Benefits	.14		
History Confidence	.13		
Expectations	.11		

Note. [a]Variables included in significant discriminant function, X^2 (2)= 18.62, $p < .001$. DF = discriminant function.

Table 19

The Average Proportion of Students at Each Participation Level within each Qualitative Level Overall and across Units.

	Unit					
	1	2	3	4	5	Average
A-Level	*a, b, c*	*a, b, e, f*	*a, b, d, e*	*a, d, e*	*a, b, e, f*	*a, b, d*
Non-responders	0/9	0/10	0/12	0/9	0/6	0/10
Occasional	5/9	7/10	7/12	7/9	3/6	6/10
Frequent	2/9	3/10	3/12	2/9	3/6	3/10
Dominant	2/9	0/10	2/12	0/9	0/6	1/10
B-Level	*c*	*a, b, c*	*b, c*	*b, c, e*		*b, c*
Non-responders	0/9	0/9	0/10	0/10	0/6	0/9
Occasional	2/9	3/9	2/10	1/10	2/6	2/9
Frequent	2/9	3/9	3/10	3/10	2/6	3/9
Dominant	4/9	3/9	5/10	6/10	2/6	4/9
C-Level	*a*	*a, c*			*a*	*a*
Non-responders	1/14	0/11	1/9	1/5	1/13	1/11
Occasional	7/14	5/11	4/9	1/5	6/13	5/11
Frequent	4/14	2/11	2/9	2/5	3/13	3/11
Dominant	2/14	4/11	2/9	1/5	3/13	2/11
D-Level		*b*	*b, c*	*b*	*b, c, d, e*	
Non-responders	4/9	3/8	4/8	5/12	5/11	4/10
Occasional	3/9	4/8	4/8	4/12	5/11	4/10
Frequent	1/9	0/8	0/8	1/12	1/11	1/10
Dominant	1/9	1/8	0/8	2/12	0/11	1/10

Note. *a* = significant difference between non-responders and occasional responders; *b* = significant difference between non-responders and frequent responders; *c* = significant difference between non-responders and dominant responders; *d* = significant difference between occasional and frequent responders; *e* = significant difference between occasional and dominant responders; *f* = significant difference between frequent and dominant responders.

Appendix B: Figures

Student Record Card

Name: _____ N Card: (□) Yes□ or No□
Unit: _____ Date: _____ IN HW: (□) Yes□ or No□
AQ HW: (□) Yes□ or No□
Voluntary Comments: Check each comment as Timely (T) or Repetitious (R).
1. T ( ) or R ( ) --

2. T ( ) or R ( ) --

3. T ( ) or R ( ) --

Over: (□) Yes□ or No□

Figure 1. Record card for recording participation and daily credit activities.

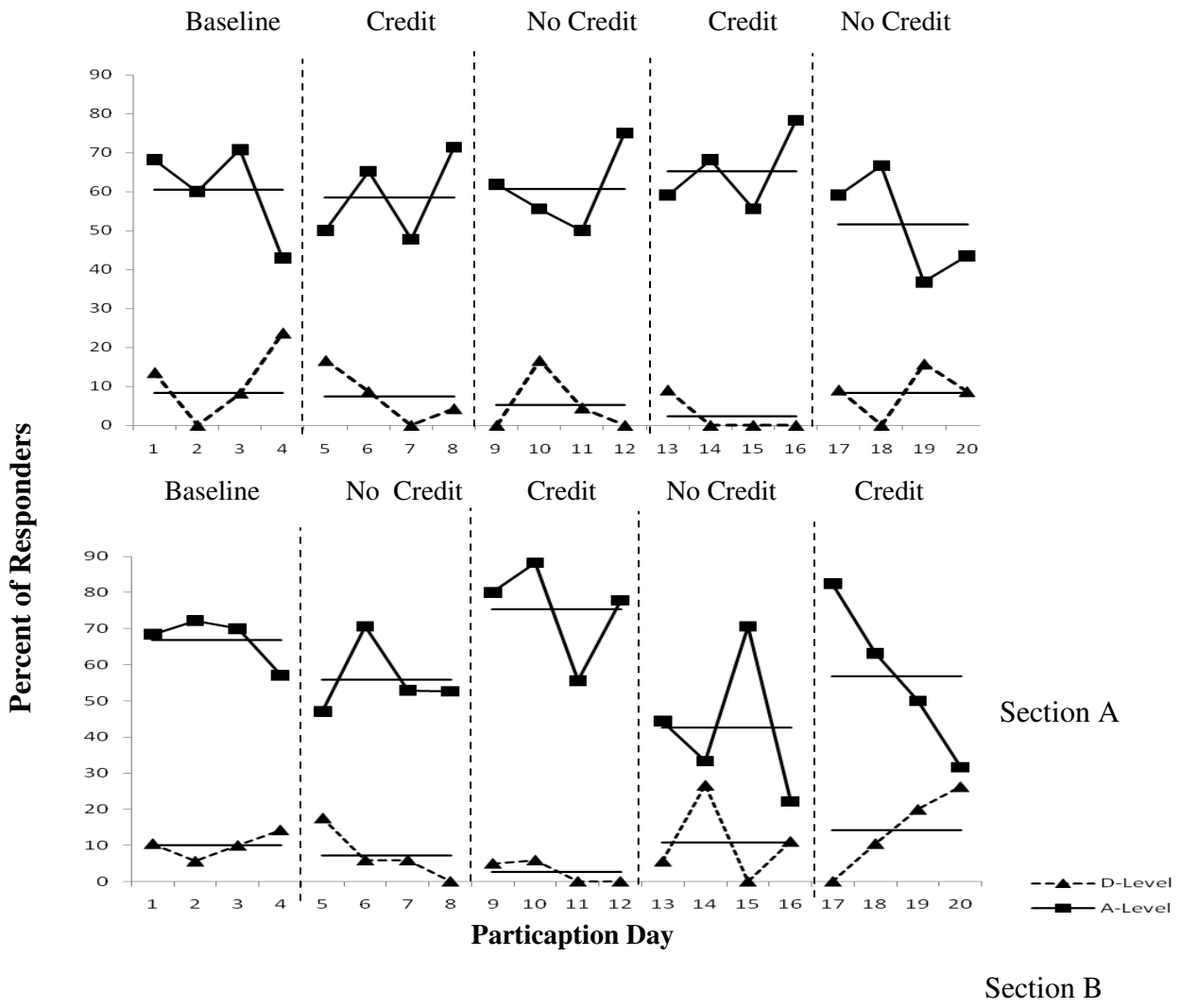

Figure 2. Percent of D-Level (67% or less productive) and A-Level responders (99 to 100% productive comments) each day.

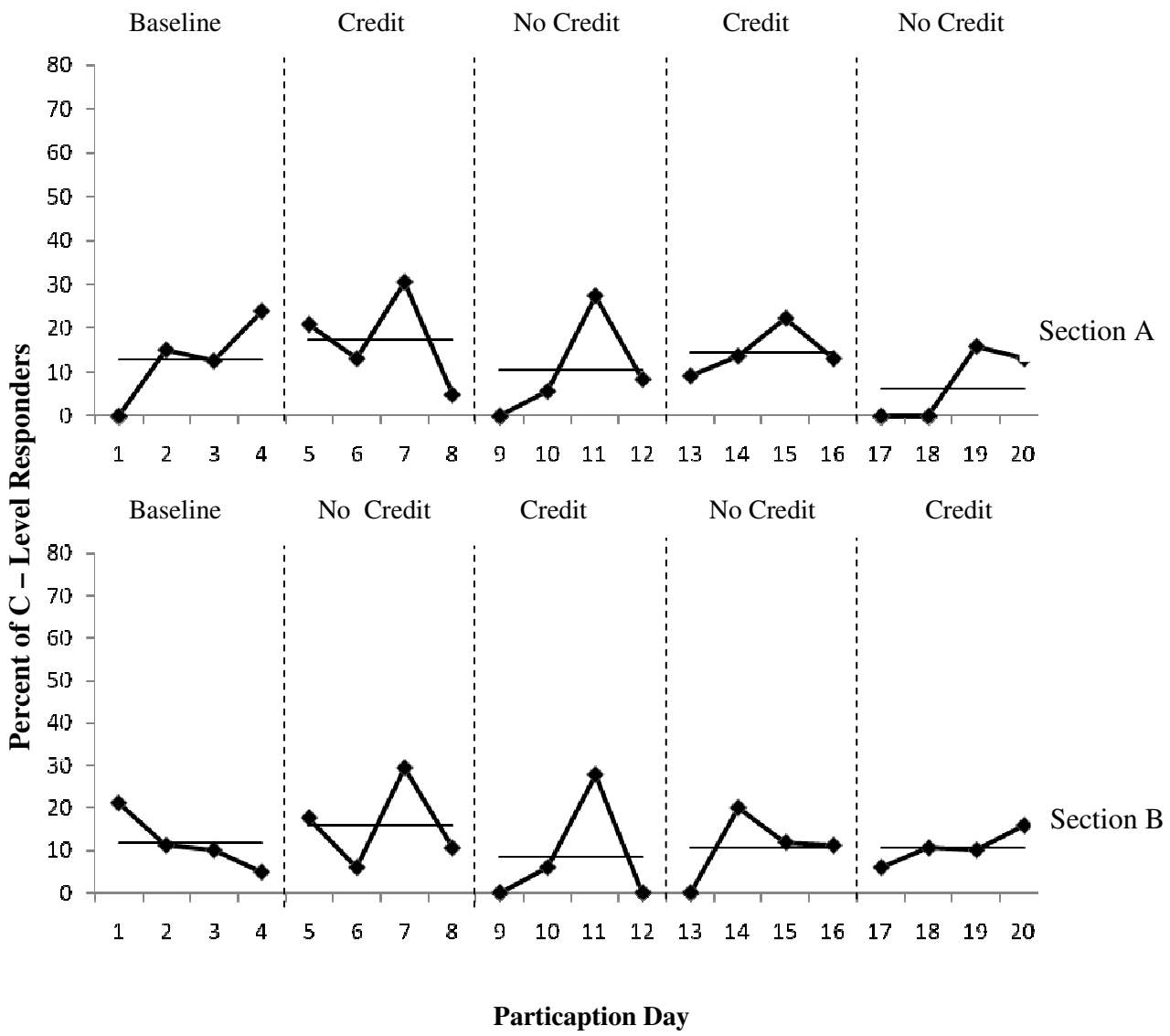

Figure 3. Percent of C-Level responders (68 - 88% productive comments) each day.

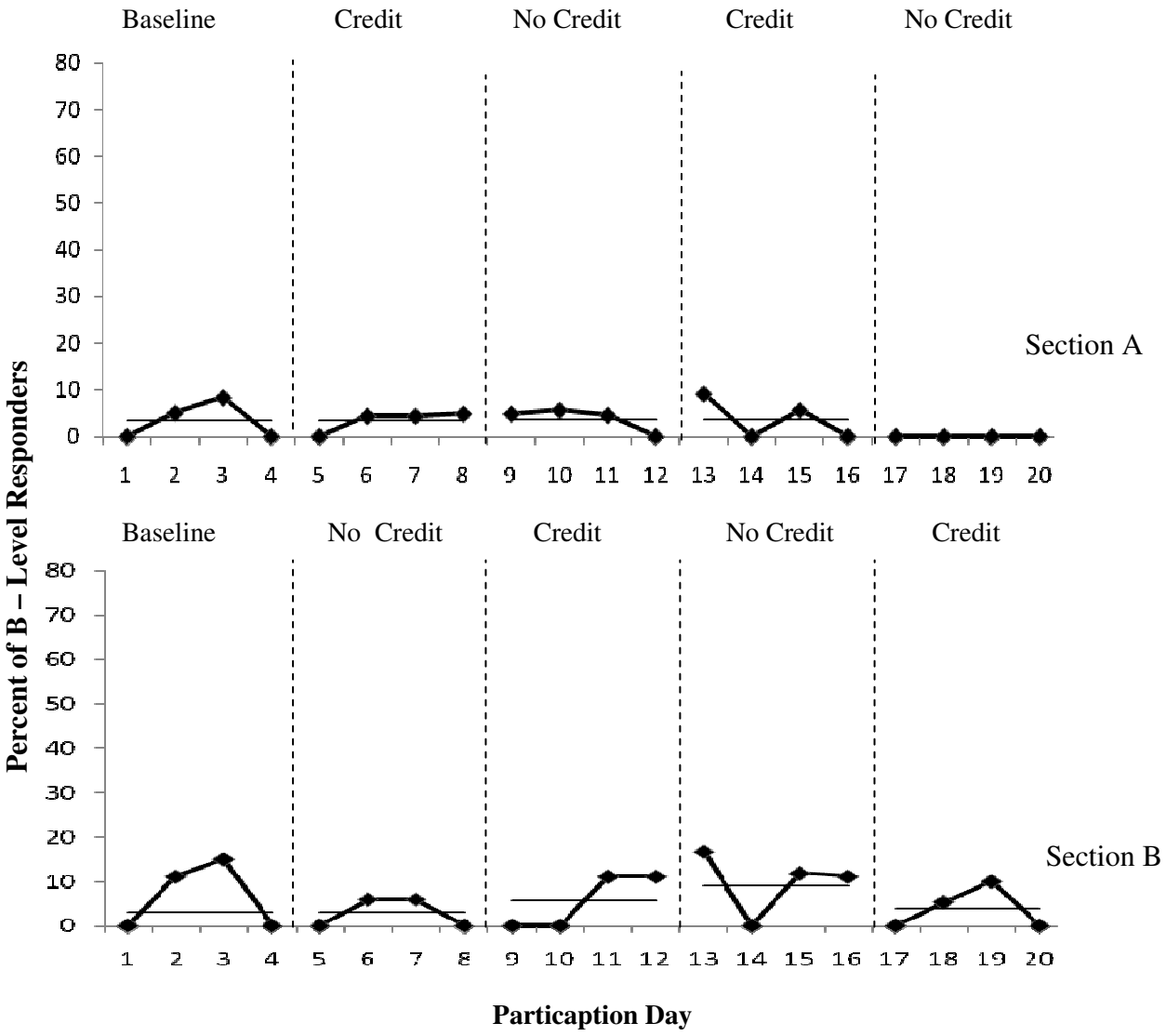

Figure 4. Percent of B-Level responders (89 - 98% productive comments) each day.

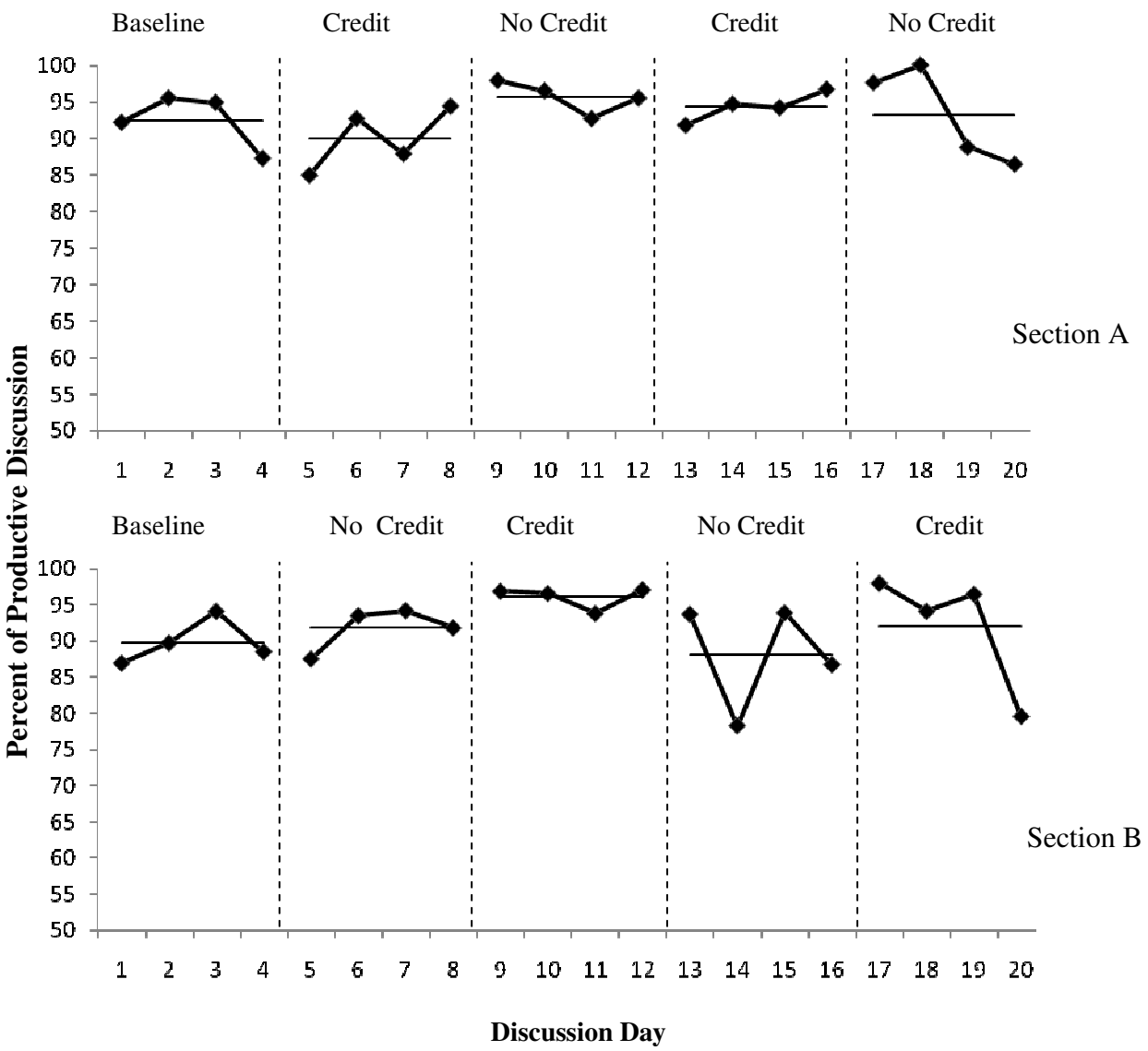

Figure 5. Percent of productive comments each discussion day.

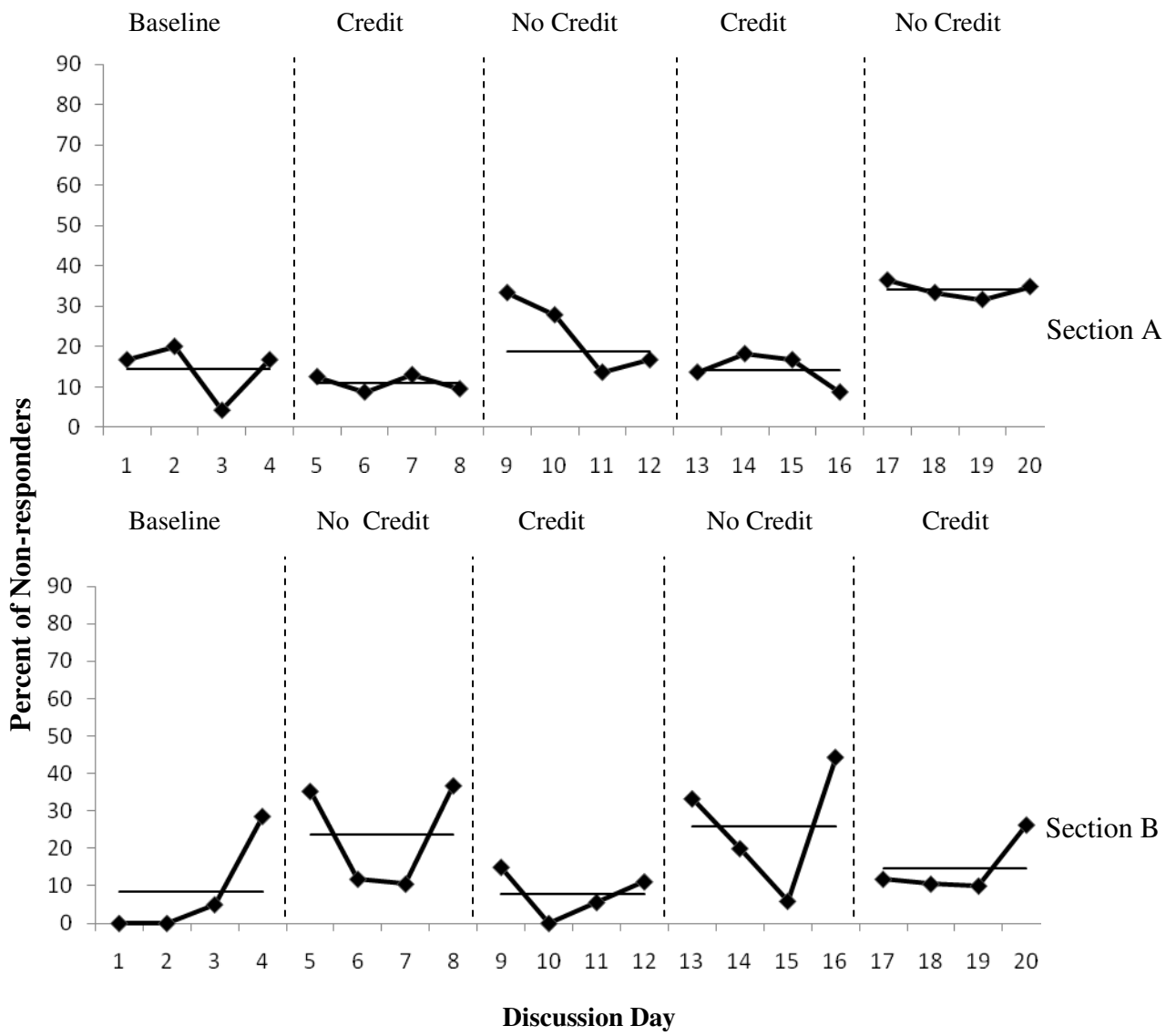

Figure 6. Percent of non-responders (0 comments) each day.

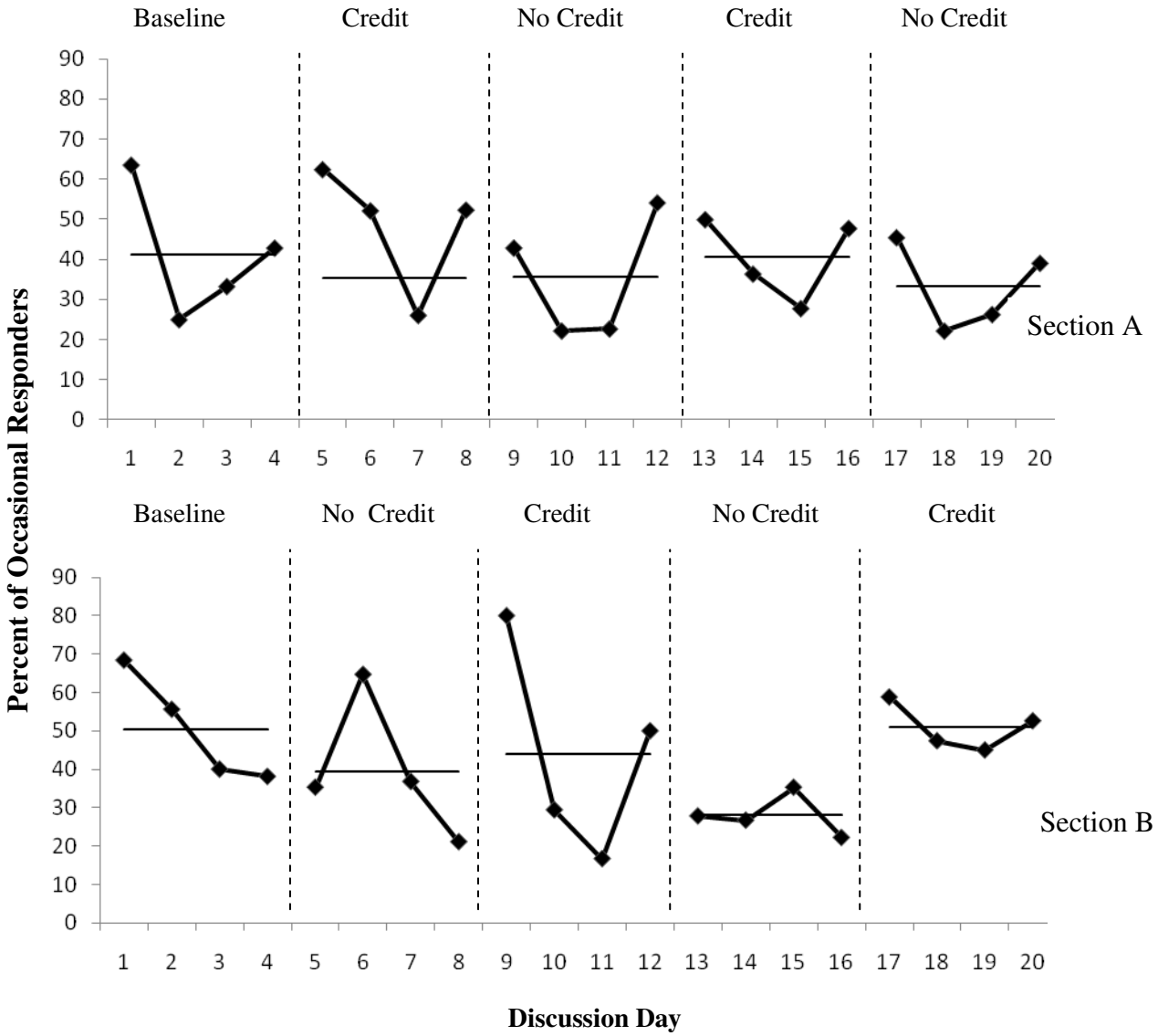

Figure 7. Percent of occasional responders (1-3 comments) each day.

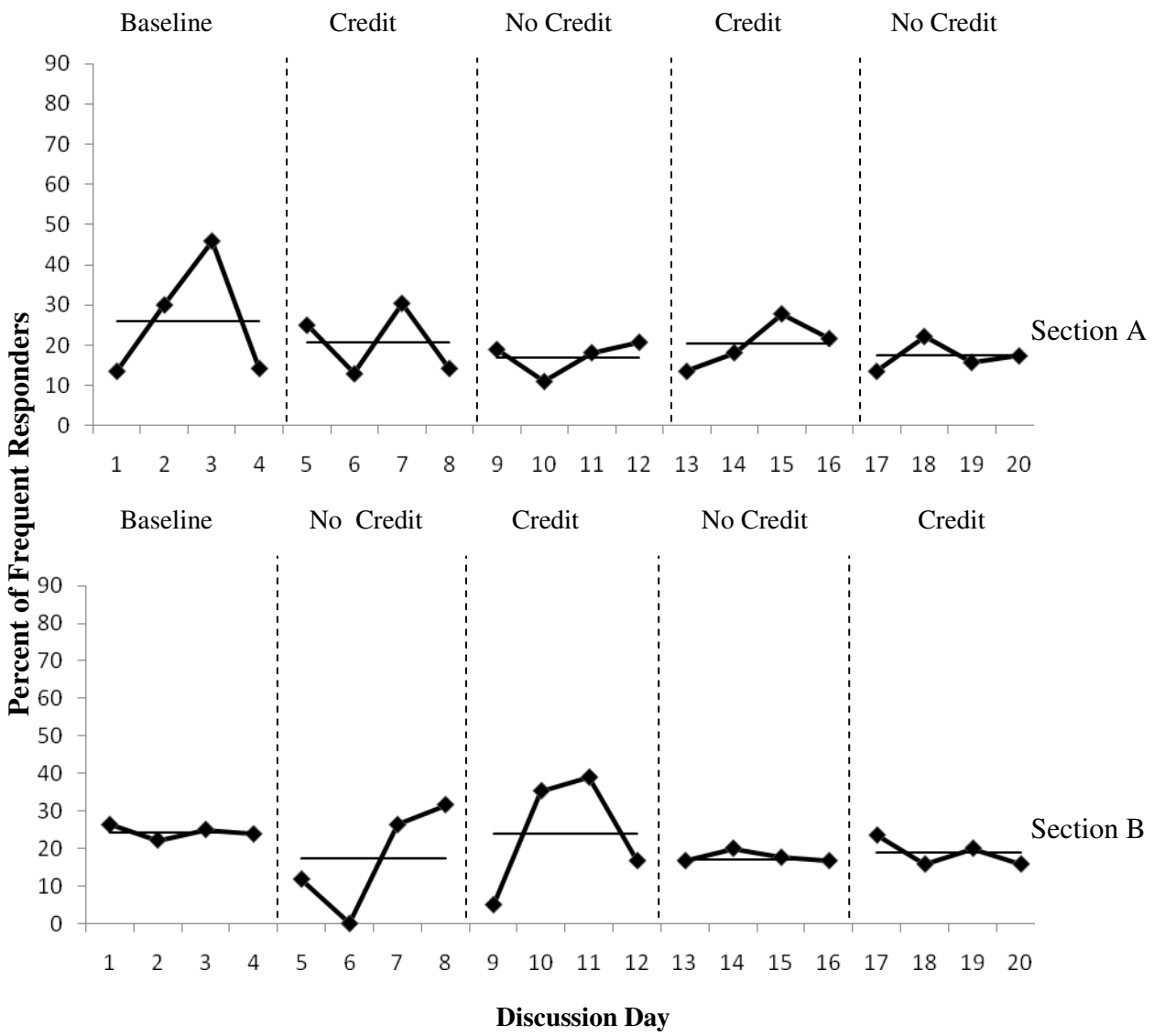

Figure 8. Percent of frequent responders (4-6 comments) each day.

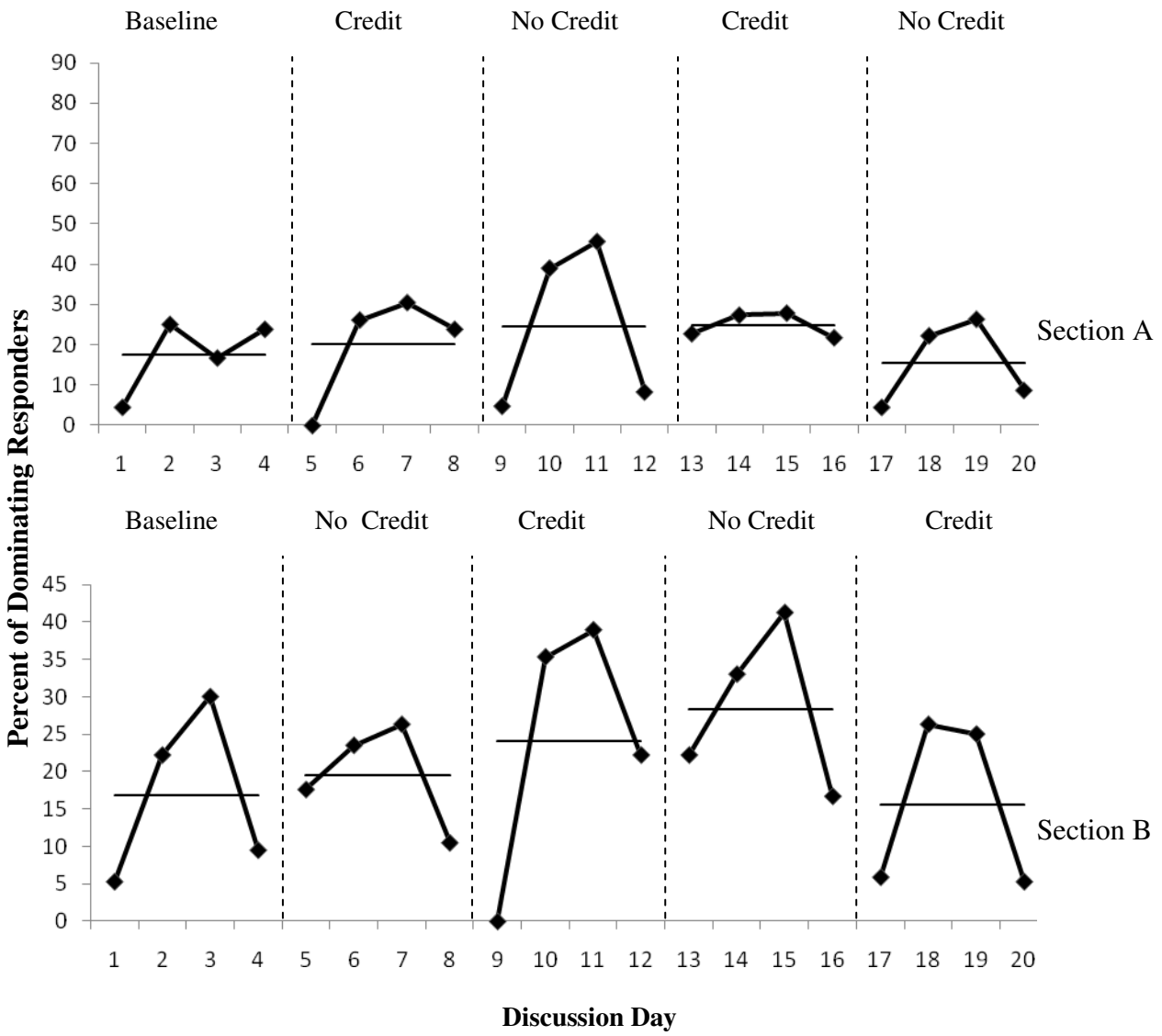

Figure 9. Percent of dominating responders (7+ comments) each day.

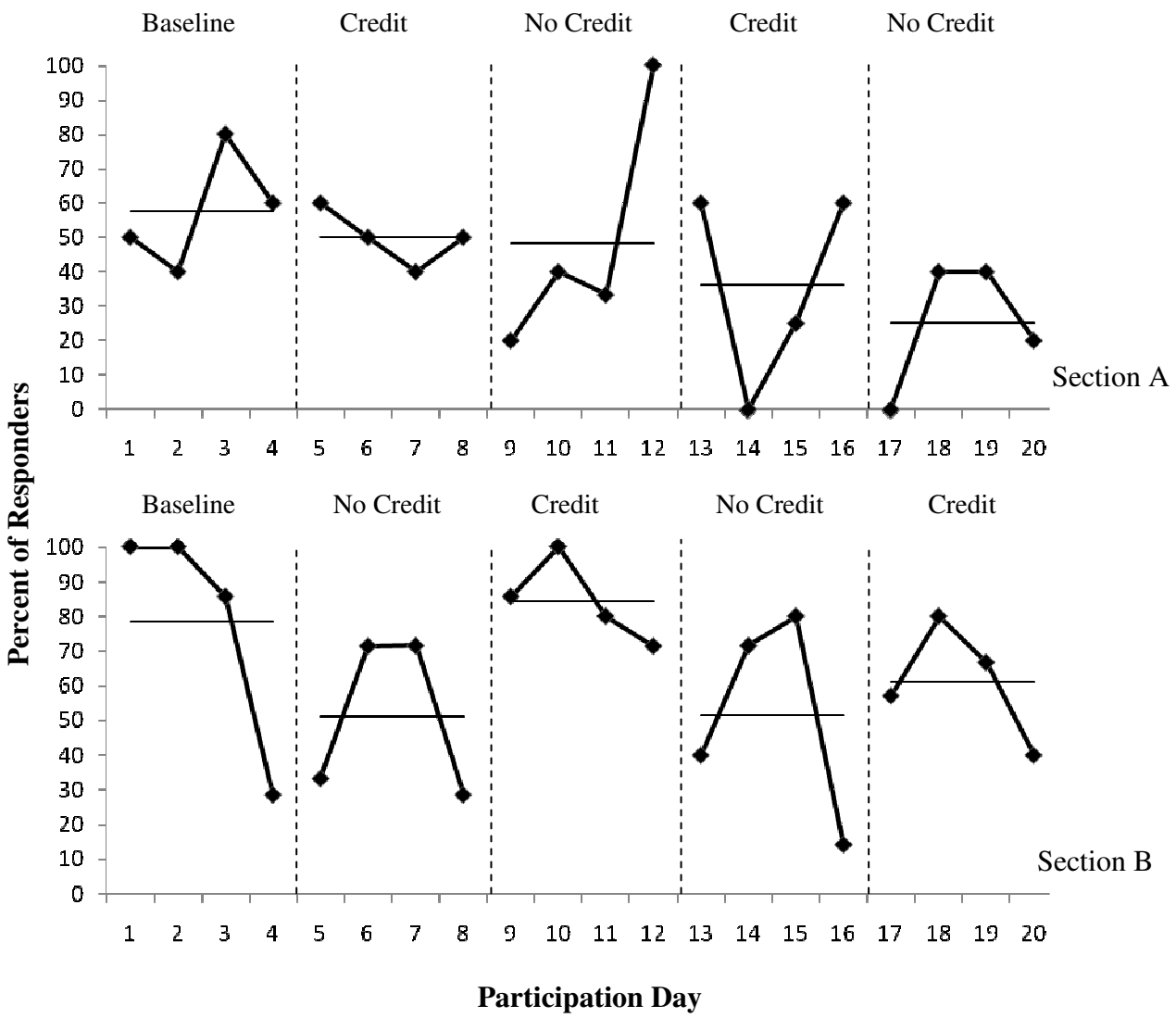

Figure 10. Percent of low-responding students in Unit 1 who participated in subsequent units.

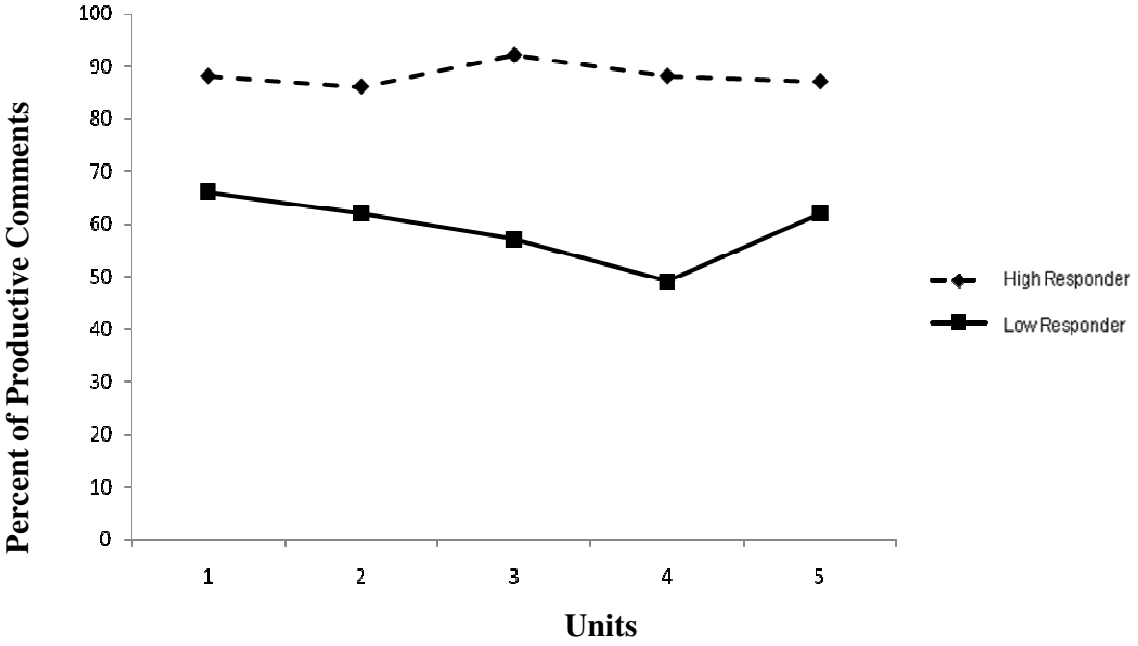

Figure 11. Overall mean percent of productive comments for high and low responders across units and sections.

Appendix C

Productive and Nonproductive Comments

Productive patterns
Asks questions or offers perspectives that extend ideas in the notes and/or class discussion
Answers a factual question that cannot be answered from information in the notes (i.e., is common knowledge)
Accurately addresses factual questions in one's own words.
Questions perspectives included in the course materials (i.e., healthy skepticism)
Locates specific passages in the notes that clarify debated issues
Asks for clarification of concepts in the notes or under discussion
Provides experiences to illustrate/explain concepts in the notes

Non-productive patterns
Poorly answers a comprehension question
Answers a comprehension question untimely
Addresses factual questions in an inaccurate or untimely manner
Gives untimely testimony
Offers "off-the-subject" jokes during class discussion.
Answers instructor questions by reading directly from the notes without demonstrating a personal understanding

Appendix D

Qualitative Assessment of Student Participation in Class Discussion

Code	Types of Student Comments in Class Discussion		
	Productive Comments		
	Comment Type	**Comment Description**	**Example**
C	Reasonably asks questions or offers perspectives that extend ideas in the course materials and/or class discussion	These comments include reasonable answers to an instructor-posed comprehension question. Comprehension questions require students to use information in the course materials to compare/contrast concepts, to develop main ideas or themes in the readings, to apply course concepts, and/or to hypothesize explanations for certain findings. The answers to these questions cannot be found directly in the course material. These comments can also include questions or statements that relate, compare/contrast, or apply course concepts or topics under discussion to something the student has read or watched such as professional sources.	*Instructor*: There is a statement here that may strike you as incorrect (teenage girls gain weight at the same rate whether they smoke or not). How can this be correct? *Student A*: Smoking may increase metabolism, but teenage girls' metabolisms are already super fast so smoking wouldn't make a difference in their rate of weight gain. (C) *Instructor:* Do girls really eat healthier than boys do? Why is that? *Student B:* I don't think so. I think it's just that girls report better eating than boys (C) Student C: I disagree. I think society really has influenced girls to eat unhealthily (C) *Student B*: I've conducted studies with a colleague of mine and we are writing a book that demonstrates that females are emotional eaters and males are not. So, females eat when they are happy or upset. (C)
A	Answers a factual question that cannot be answered from information in the notes (i.e., is common knowledge).	These responses are instructor elicited. They are responses to questions about information that is beyond the scope of the course, yet pertains to the topic being discussed.	*Instructor:* What is the procedure for random selection? *Student:* You select participants at random rather than including them in a study because they are in a certain class. (A)

F	Accurately addresses factual questions in one's own words.	The student accurately answers an instructor-posed factual question in his or her own words rather than reading verbatim from the course materials. The student may reference the course materials during his or her response, but does not simply read word for word from the course materials. A factual question is any question that can be answered directly from the course materials.	*Instructor:* Overall, what is the peak year for drug use? *Student*: 1976, but overall drug use has gone down (F)
S	Questions perspectives included in the readings.	The student questions the accuracy or logic of information in the readings (instructor notes and articles) by offering logical reasoning or empirical evidence for the questioning. The comment does not simply state that a claim is incorrect.	*Student*: I think the number of children in the schools that is in the notes may be inaccurate due to the influx of immigrants and people who migrate to work on farms in the U.S. (S)
L	Locates specific passages in the readings that clarify debated issues.	While an issue is under debate in class discussion, the student clarifies the issue by pointing to a passage that directly addresses the issue.	*Student 1*: I don't think females eat better than males at all. That must be wrong [This student's comment would be classified under (D)] *Student 2:* The notes do acknowledge that the finding may be due to dieting practices that can lead to unhealthy dieting. (L)
C	Asks for clarification of concepts in the readings or class discussion.	The student asks a question that directly references a concept found in the course materials or concepts in class discussion. The main goal of the question is to achieve clarity (i.e., the student is initially confused) and not to point out flaws in another's logic. The optimal way to ask for clarification is for students to seek confirmation that their current understanding of a concept (e.g., "Does this concept mean . . . ?") is on target.	*Student:* I still don't understand the slide that looks at the UK doctors who smoke. The place where it says 10 years really confuses me. Can you explain it a little more? (C)
T	Provides experiences to illustrate/explain concepts in the readings.	The student may describe an experience that illustrates or explains concepts in the readings. These examples may be instructor requested.	*Instructor*: Do you faithfully follow these safe driving practices? *Student*: I may not always put on my seatbelt, but I've gotten several tickets. In this example the student demonstrates that she has some understanding of information in the notes (i.e., the driving practices referenced by the instructor).(T)
O	Other productive comments	Productive comments that cannot be categorized in another category.	

Code	Nonproductive Comments		
	Comment Type	**Comment Description**	**Example**
XC	Untimely or poorly answers a comprehension question.	The student answers a comprehension question already answered by another student without adding anything new to the student's previous comment. Many comprehension questions do not have one answer, leaving students to speculate about relationships between concepts included in the readings. However, if a student's answer to a comprehension question only remotely relates to that question or contradicts information found in the course materials without the student's citing evidence from professional sources to support the disagreement with course materials, the student comment should be classified as a poor answer to the comprehension question.	*Instructor:* Should teachers be screened for fitness? *Student 1:* Teachers should definitely be a role model for their students especially if they are gym teachers. (C) *Student 2*: What about a teacher who is an excellent teacher but is obese? Obesity tends to run in families and can be hard to fight despite diet and exercise. (C) *Student 3*: I think it's definitely important for gym teachers. (XC) *Instructor*: What is the difference between HDL and LDL? *Student:* HDL and LDL are cholesterol and should be minimized. (XC)
XF	Addresses factual questions in an inaccurate or untimely manner.	Inaccurately answers a factual question or provides an answer that has already been provided.	*Instructor*: What are some reasons smoking contributes to heart disease? *Student 1:* Smoking decreases LDL which is the good cholesterol and raises HDL which is the bad cholesterol (F) *Student 2*: Smoking can cause pits in the arteries that lead to atherosclerosis (F) *Student 3*: Smoking increases bad cholesterol (XF)
XT	Gives untimely testimonial	Provides testimony that is only tangentially relates to issues under discussion or provides redundant testimony. Testimony is redundant when the student provides it after the instructor has requested that no more testimonies be given or has otherwise suggested that students move on to another issue.	*Instructor*: So, smoking increases the probability of neuromuscular diseases? *Student 1*: No, not all of them. It actually reduces symptoms of Tourettes. (F) *Student 2*: I think that smoking can reduce or alleviate migraines too, at least that's what my doctor says (T) *Student 3*: My great-grandmother had Tourette's. She also had cancer. It's an interesting story really. She grew up in a time when Tourette's wasn't really understood which caused a lot of emotional problems for her. (XT)

J	Offers "off-the-subject" jokes during class discussion.	The student offers jokes that are not related to the content being discussed.	*Instructor*: So women report more health problems but live longer. How can you explain that? *Student A*: Hereditary advantages (F) *Student B*: Because they are more likely to go to the doctor (F) *Student C*: They have a stronger social support. (F) *Student D*: Because they have fun after their husbands are gone (J)
R	Answers instructor questions by reading directly from the notes without demonstrating a personal understanding.	The student spends the majority of her response looking at the course materials and reading from the course materials, verbatim.	*Instructor*: So, what does this slide illustrate? *Student*: A UK study of male doctors born between 1900 and 1930 shows that stopping smoking by age 30 essentially eliminates the risk of premature death from smoking and stopping by age 50 cuts the risk in half. (R)
D	Expresses disagreement in a nonproductive manner.	The students' attack another student's ideas, using words such as "stupid," "ridiculous," or "uninformed." The student expresses disagreement with others' ideas or with course concepts without providing logical reasoning for that disagreement.	*Student:* Daily tobacco use seems really low, are you sure that's right? (D)
XO	Other nonproductive comments	These comments include comments that are deemed nonproductive yet cannot be correctly classified under a specified category.	

Appendix E

Rating Form

Instructor: _____ Section: ___ Unit:_____ Day: ____ Date: _____ Rater: _____

	Comment												Key
Name	1	2	3	4	5	6	7	8	9	10	11	12	Productive C – COMP A – ANWER F – FACT S – SKEP L – LOCATE FY – CLRFY T – TEST O – OTHER Nonproductive XC - XCOMP XF – XFACT XT – XTEST J – JOKE R – READ D – DISS XO – OTHER

Appendix F

Practice Script

Instructor: What is the best source of information?

Anna: Professional journals

Instructor: Why do you think professional journals are more credible than the others?

Tina: Because there is a review process to ensure quality so not just any thing can get published.

Instructor: What is the weakest source of information?

Jack: I think it would probably be trade-market books because they have absolutely no scientific backing. They can just be made up and if proper advertising is done, they can be sold.

Instructor: What are some examples of trade-market books?

Anna: The Atkins diet?

Instructor: What does it mean to say that two things correlate?

Manny: That they go together somehow. That they are related.

Instructor: Right, so if I say that variables have a relationship of positive 1, what does that say about the relationship between the two variables?

Ronald: That as one goes up the other one goes up and that they have perfect predictability.

Instructor: What does it mean to say that heart rate and amount of activity have a zero correlation?

Edith: That your dead.

Instructor: Not quiet, can anyone help Edith out?

Manny: It means that there is no predictable relationship between heart rate and amount of activity.

Instructor: What are some qualities of an experiment?

Ronald: Random assignment and the manipulation of some variable.

Instructor: Does anyone know the procedure for random assignment?

Jack: You select people randomly from a group.

Edith: I've conducted experiments in schools and we match participants on gender and socioeconomic status because we can't just randomly assign them to conditions. It does help to control for extraneous variables though.

Instructor: That's a good point Edith. You're talking about random assignment of participants to conditions, not random selection of participants. But, it is important to randomly assign participants to conditions as well.

Ronald: We've talked in my previous classrooms about stratified sampling where you randomly select participants from certain groups, like every state, in order to get a representative sample which seems to be a good way to randomly select rather than just pulling random numbers.

Instructor: What are some differences in boys and girls when it comes to exercise?

Manny: (looking in book) Boys prefer competitive sports whereas girls are more likely than boys to prefer noncompetitive physical activities.

Instructor: Can you put that in your own words?

Manny: Boys play competitive sports and girls like activities that are not competitive like aerobic exercise.

Edith: Yeah, all the girls I know go for runs and all the boys I know prefer to play a sport.

Instructor: Why do you think boys prefer more competitive activities?

Anna: Because they have more testosterone and are naturally more aggressive.

Edith: Boys' sports are more publicized than girls' sports are in the media.

Manny: I think that this may be changing (mumble mumble)….nevermind.

Ronald: Another reason might be because girls grow up watching their moms and boys grow up wanting to be like dads and dads are more likely to be active in a competitive sense.

Tina: I disagree, my daughter will not do what I'm doing. She always wants to be outside and spends a lot of time around her father.

Instructor: How much more likely are girls to have an eating disorder than boys?

Jack: Twice as likely?

Instructor: Not quiet.

Edith: 9 to 1 to 6 to 1

Instructor: What about men with eating disorders? Is it really less?

Ronald: I don't think so. I think that we overlook men who are eating unhealthily to achieve a certain weight and it's not socially acceptable for them to have an eating disorder.

Instructor: What are some sports with high incidents of eating disorders for males?

Jack: Wrestling, because they are trying to achieve a certain weight.

Ronald: I was a wrestler and not eating to achieve a certain weight started for me in 7th grade.

Jack: Ya, it was 7th grade for me too.

Instructor: What are some other ways that boys and girls differ when it comes to health practices?

Jack: Girls report more stress and are safer drivers.

Edith: Boys smoke more than girls

Jack: Girls eat better than boys.

Appendix G: Students' Under- and Over-Reporting of Class Participation during Non-Credit vs. Credit Units

Sec. A	Non-Credit Units								Credit Units							
Student #	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
1	=	=	=	=	=	<	=	=	=	=	=	=	=	<	=	=
2	=		=	>p	=	<n	=	=	=	>p	<	=	>	=	<	<
3		<	<	<n	<		<n	<n	<	<	<	<	<	<		<
4	>	<	=	<	>	<	<	>	>	=	>	<	>	<	=	=
5	=		<	=	=	>p	<n	=	=	>p	=		<		<	<
6	=	<n		<n		=	=	<n	=	=	<	=	=	=	=	> p
7	<		<	<	=	<	<	<	<	<	<	<	<	<	<	<
8	=		<	>p	=		<	<	<	<	<	<		<		<
9		<n	=	<	=		=	=	=		=	=	=	=	=	=
10	<	<	<	=	<	<	<	<	<	<	<	<	<	<	<	<
11	=	<	<	=	=	<n		=	<	<	>	<	<	=		=
12	=	<	<	=	<			<	>p	<	<		<	<	<	<
13	=	>p	<	=	=	<		>p	=	<	<	=		<	=	<
14	=		<	=		=		<n	=	=	<	>	=	=		<
15	=	=	=	=	=			=	=	=	=	=	=	=		=
16	=	=	=	=	<n	=	=	=	=	=	<	=	=	=	=	=
17	<n	<n	<	<	<n	<	=		<	<	<		=	<	<	<
18	>p		<	=	=	=	<n	=	>p	=	<	=	=	=	=	
19	<n	<	<	<	=	<	<	<n	<	<		<	<	<	<	<
20	=	=		<	=	=	=	=	=	=	=	=	=	=	=	=
21	=	<	<	<	<n	<	=	<n	=	<	<	=	<	<n	=	=
22	<	<	=	=	=			=	=	=	<	<	=	=		=
23		=	=	=	=	=	=	=	=	=	<	=	=		<	=
24	=	<	<	=	<	<	<	<	=	=	=	=	<	<	<	=

Sec. B	Non-Credit Units								Credit Units							
Student #	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
25	=	<	=	=		=	=	=	=	=		=				
26	=	<	<n	=	=	<	=	=	=	=		=	=	=	>	=
27	<	=	<n	=		=		=	=	>	=	>	>p	>p	>p	
28	=	=	<n	=	=	=	=	=	=	<	<	=	=	=	=	=
29	=	=	=			<	=	=	<n	<	=	<	=	=	=	=
30	>	=	<	<	<n	<	<		=	<	<	<	=	<	=	=
31	<	<	<	<	<	<	<	<	=	<	<	<	<	<	<	<
32		=	=	>f	=		<	>f	=	=	<	>p		=	<	<
33	<	<	<	<	<		<	<	=	<	<	<	<	<	<	< n
34	=	=	<	=	<	=	<	=	>p	=	=	=	>p	=	=	=
35		<	<	=	=	<	<	=	=	=	<	<	<	=	<	< n
36	<			<	<n		<	<n	=		<		>f	>	<	<
37	=	=	=	=	=	=	=	=	=	=	=	=		=	=	=
38	=	>p		=	=	=			=		=	=	=		<	> p
39	=	<n	<	>p	<	<		<n	=	<	<	=	<	<	<	=
40	<	=		<	<			=		<	<	=	>	>		=
41	<	<	<		<		<n	<n	=	<	<			<	<	=
42		<	=	<	<	<	<	<	=		<		<	<	<n	>
43	<			<	<	=	<	<	<			<	<	<	<	<
44	<			<	=		<		=	<	<	<	=	<	<	<
45			<	<	=	<	<	<	=	<	<	<	<	<	=	<
Total =	22	13	13	21	22	13	14	20	31	17	11	20	17	17	15	20

Total >	2	0	0	0	1	0	0	0	1	1	2	2	3	2	1	1
Total >p	1	2	0	2	0	1	0	1	3	2	0	0	2	1	0	1
Total >f	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
Total <	11	16	21	15	12	17	17	10	8	20	28	16	16	20	19	17
Total <n	2	1	3	2	5	2	4	8	1	0	0	0	0	1	1	2

Note. > student over-reporting; < student under-reporting; = agreement between student and observer; >p student over-reporting resulted in partial credit unjustly awarded; >f student over-reporting resulted in full credit unjustly awarded; <n student under-reporting resulted in not receiving credit.

Appendix H: Participation Survey

Following each item stem is the mean and standard deviation for the total sample. The possible range for each item is 1 (option e) to 5 (option a). Note that option a for each item generally indicates a more favorable nature of participation (e.g., greater comfort participating, higher previous or anticipated levels of participation). The percent of students that selected each item option is also provided in parenthesis.

1. Which of the following best describes your previous pattern of class participation in college courses?
    a. participating several times a day in most class discussions
    b. participating once or twice a day in most class discussions
    c. participating once or twice a day in about 50% of the class discussions
    d. participating infrequently in class discussions
    e. never participating in class discussions

2. Which of the following best describes your expectations for participating in discussions in the 210 course?
    a. participating several times a day in most discussions
    b. participating once or twice a day in most discussions
    c. participating once or twice a day in about 50% of the discussions
    d. participating infrequently in discussions
    e. never participating in discussions

3. Which of the following best describes your feelings about participating in discussions in classes as large as the 210 course?
    a. consistently feel comfortable when participating
    b. generally feel comfortable when participating
    c. feel comfortable about half the time when participating
    d. generally feel uncomfortable when participating
    e. consistently feel uncomfortable when participating

4. Which of the following best describes how you feel when called on to participate in class discussion?
    a. extremely confident
    b. somewhat confident
    c. neutral
    d. somewhat uncomfortable
    e. extremely uncomfortable

5. Which of the following best describes your perspective of the relevance of your comments in class discussion?
    a. Your comments are almost always relevant.
    b. Your comments are generally relevant.
    c. Your comments are relevant about half the time.
    d. Your comments are seldom relevant.
    e. Your comments are almost never relevant.

6. Which of the following most accurately describes your typical level of preparation for class discussion in past courses?
    a. completed all of the homework related to the topic to be discussed plus done some additional investigation of the topic
    b. completed all of the homework related to the topic to be discussed
    c. completed most of the homework related to the topic to be discussed
    d. completed little of the homework related to the topic to be discussed
    e. completed none of the homework related to the topic to be discussed

7. What is your attitude about student responsibility for participating in class discussion when the instructor asks students to volunteer responses to instructor questions?
    a. Students are totally responsible for volunteering comments under those circumstances.
    b. Students bear most of the responsibility for volunteering comments under those circumstances.
    c. Students have marginal responsibility for volunteering comments under those circumstances.
    d. Students have little responsibility for volunteering comments under those circumstances.
    e. Students have no responsibility for volunteering comments under those circumstances.

8. What is your attitude about earning course credit for participating in class discussion?
    a. Participation should be the most heavily weighted contributor to your grade.
    b. Participation should be substantially weighted in the computation of your grade.
    c. Participation should be moderately weighted in the computation of your grade.
    d. Participation should be minimally weighted in the computation of your grade.
    e. Participation should not be considered in the computation of your grade.

9. How do you feel when an instructor indicates at the beginning of a course that students will be expected to participate in class discussion?
    a. extremely positive
    b. generally positive
    c. neutral
    d. generally negative
    e. extremely negative

10. How do you feel when an instructor indicates at the beginning of a course that class discussion is welcomed but optional?
    a. extremely positive
    b. generally positive
    c. neutral
    d. generally negative
    e. extremely negative

11. What would be the relative likelihood of your asking a question versus answering a question in class discussion?
    a. much more likely to ask a question
    b. somewhat more likely to ask a question
    c. about equally likely to ask or answer a question
    d. somewhat more likely to answer a question
    e. much more likely to answer a question

12. How would you describe the general effect of your participating in discussion on your learning in a course?
    a. learn best when participating heavily in discussion
    b. learn best when participating regularly in discussion
    c. learn best when participating periodically
    d. learn best when participating infrequently in discussion
    e. learn best when never participating in discussion

13. How do you think that your keeping a record of your participation in class discussion would affect your concentration on the discussion?
    a. would greatly increase your concentration on the discussion
    b. would generally contribute to your concentration on the discussion
    c. would have an uncertain effect on your concentration on the discussion
    d. would generally detract from your concentration on the discussion
    e. would greatly detract from your concentration on the discussion

14. Which of the following class formats (discussion versus lecture) do you prefer in courses you take?
    a. all discussion
    b. mainly discussion but some lecture
    c. a balance between discussion and lecture
    d. mainly lecture but some discussion
    e. all lecture

15. What do you see as the relationship between the quantity and quality of class discussion?
    a. Quantity consistently contributes to quality.
    b. Quantity is more likely to contribute to quality than detract from it.
    c. Quantity and quality are unrelated.
    d. Quantity is more likely to detract from quality than contribute to it.
    e. Quantity consistently detracts from quality.

16. How do you typically respond when an instructor poses a question for class discussion?
    a. quickly speak up
    b. speak up after a short delay
    c. speak up but with considerable hesitancy
    d. speak up only if no one else speaks up
    e. not speak up even if no one else speaks up

17. How do you feel about a discussion format in which students volunteer comments rather than being called on by the instructor?
    a. greatly prefer volunteering comments rather than being called on
    b. somewhat prefer volunteering comments rather than being called on
    c. equally comfortable with volunteering and being called on
    d. somewhat prefer being called on rather than volunteering
    e. greatly prefer being called on rather than volunteering

18. Who is responsible for a *high level of student participation* in class discussion?
    a. exclusively the students
    b. primarily the students
    c. shared equally between the students and the instructor
    d. primarily the instructor
    e. exclusively the instructor

19. Who is responsible for *very limited student participation* in class discussion?
    a. exclusively the students
    b. primarily the students
    c. shared equally between the students and the instructor
    d. primarily the instructor
    e. exclusively the instructor

20. How would a class with frequent discussion affect your evaluation of the course?
    a. greatly increase your evaluation of the course
    b. generally increase your evaluation of the course
    c. have little effect on your evaluation of the course
    d. generally decrease your evaluation of the course
    e. greatly decrease your evaluation of the course

21. What effect does frequent discussion by other students have on your concentration in class?
    a. greatly increases your concentration
    b. moderately increases your concentration
    c. minimally affects your concentration
    d. moderately decreases your concentration
    e. greatly decreases your concentration

22. What effect does your personal participation in class discussion have on your concentration in class?
    a. greatly increases your concentration
    b. moderately increases your concentration
    c. minimally affects your concentration
    d. moderately decreases your concentration
    e. greatly decreases your concentration

23. What effect does the option of volunteering comments whenever you wish have on your concentration in class?
    a. greatly increases your concentration
    b. moderately increases your concentration
    c. minimally affects your concentration
    d. moderately decreases your concentration
    e. greatly decreases your concentration

24. How does the possibility that you might be called on to respond to an instructor question affect your concentration in class?
    a. greatly increases your concentration
    b. moderately increases your concentration
    c. minimally affects your concentration
    d. moderately decreases your concentration
    e. greatly decreases your concentration

25. How does frequent discussion in the class as a whole affect your enjoyment of a class?
    a. makes the class much more enjoyable
    b. makes the class somewhat more enjoyable
    c. doesn't affect your enjoyment one way or the other
    d. makes the class somewhat less enjoyable
    e. makes the class much less enjoyable

26. How do you feel toward students who frequently comment in class discussion?
    a. You greatly appreciate their frequent participation.
    b. You generally appreciate their frequent participation.
    c. You feel neutral toward their frequent participation.
    d. You are generally annoyed by their frequent participation.
    e. You are greatly annoyed by their frequent participation.

27. What are your academic expectations of students who frequently participate in class?
    a. You expect them to do well in the course.
    b. You expect them to do somewhat better than average in the course.
    c. You expect their frequent contributions to be unrelated to their performance in the course.
    d. You expect them to do somewhat worse than average in the course.
    e. You expect them to do poorly in the course.

28. Some students like to be knowledgeable about a course topic before contributing to class discussion on that topic. How do you feel about this issue?
    a. You have the strongest inclination to comment on topics about which you have the most knowledge.
    b. You are moderately inclined to comment on topics about which you have the most knowledge.
    c. Your knowledge about topics has little effect on your tendency to comment on those topics.
    d. You feel somewhat less need to comment on topics about which you have the most knowledge.
    e. You feel the least need to comment on topics about which you have the most knowledge.

29. To what degree does student sharing of personal experiences in class discussion contribute to the quality of the discussion?
    a. greatly heightens the quality of class discussion
    b. moderately heightens the quality of class discussion
    c. has a neutral impact on the quality of class discussion
    d. moderately diminishes the quality of class discussion
    e. greatly diminishes the quality of class discussion

30. How do you typically respond when no one else is responding to a teacher question?
    a. Attempt to answer the question when no one else is responding.
    b. Wait until the silence has become somewhat uncomfortable to you before attempting to answer the question.
    c. Wait until the silence has become extremely uncomfortable to you before attempting to answer the question.
    d. Respond only if the instructor calls on you to answer the question.
    e. Decline to respond to the question even if the instructor calls on you.

31. Do you believe you have insights about course concepts that would benefit your peers if you shared them in class?
    a. definitely “yes”
    b. generally “yes”
    c. uncertain
    d. generally “no”
    e. definitely “no”

32. What effect do long pauses between teacher questions and student responses have on your desire to participate in class discussion?
    a. greatly increases your desire to participate
    b. moderately increases your desire to participate
    c. minimally affects your desire to participate
    d. moderately decreases your desire to participate
    e. greatly decreases your desire to participate

33. How would teachers in your past college courses most likely characterize your level of participation in class discussion?
    a. the most talkative student in class
    b. among the more talkative students in class
    c. talkative to an average level
    d. among the less talkative students in class
    e. the least talkative student in class

34. How would you characterize teacher views regarding the inclusion of class discussion in student grades in your past college courses?
    a. Participation is the most important part of a student’s grade.
    b. Participation is among the more important contributors to a student’s grade.
    c. Participation is on par with several other contributors to a student’s grade.
    d. Participation is among the less important contributors to a student’s grade.
    e. Participation is not included in a student’s grade.

35. In comparison to other classes you are taking this semester, what expectation do you have for your participation in 210 class discussion?
    a. more participation in 210 discussion than in any other class
    b. more participation in 210 discussion than in most other classes
    c. about the same level of participation in 210 discussion as in other classes
    d. less participation in 210 discussion than in most other courses
    e. less participation in 210 discussion than in any other course

36. Which of the following best expresses your view of the long-term value of learning to express one's views in public?
    a. Learning to express one's views in public is among the most important skills one can develop in school.
    b. Learning to express one's views in public is among the more important skills one can develop in school.
    c. Learning to express one's views in public is an important skill but certainly not among the more important skills one can develop in school.
    d. Learning to express one's views in public is among the lesser skills one can develop in school.
    e. Learning to express one's view in public is among the least important skills one can develop in school.

37. Your interpretation of how teachers feel about class discussion is best reflected in which of the following claims?
    a. Most teachers strongly value class discussion.
    b. Most teachers moderately value class discussion.
    c. Most teachers are neutral toward class discussion.
    d. Most teachers moderately devalue class discussion.
    e. Most teachers strongly devalue class discussion.

38. How would most of your high school teachers likely describe your participation in class?
    a. extremely verbal in class
    b. generally verbal in class
    c. occasionally verbal in class
    d. generally quiet in class
    e. extremely quiet in class

39. Many teachers try to stimulate class discussion by asking questions. Which of the following best expresses your view of most teacher questions?
    a. Most are highly challenging.
    b. Most are moderately challenging.
    c. Most are routine in nature.
    d. Most provide little challenge.
    e. Few provide any challenge at all.

40. At the completion of a class session in which you participated frequently, how would you most likely feel about possible classmate reaction to your comments?
    a. You would feel your classmates strongly valued your comments.
    b. You would feel your classmates moderately valued your comments.
    c. You would feel that your classmates were neutral toward your comments.
    d. You would feel that your classmates moderately devalued your comments.
    e. You would feel that your classmates strongly devalued your comments.

41. Which of the following best represents your ability to judge the relevance of your comments in class discussion?
    a. You can determine whether a comment will be relevant even before you make the comment.
    b. You have your first sense of whether a comment is relevant as you are making the comment.
    c. You can tell whether a comment is relevant only by the instructor's reaction to the comment.
    d. You can only judge the relevance of your comment when you have time to reflect on it after class.
    e. You never really have a sense of whether your comment was relevant.

42. How would frequent participation in college classes likely affect your grades in those courses?
    a. Consistently raise your grades.
    b. Generally raise your grades.
    c. Have little effect on your grades.
    d. Generally lower your grades.
    e. Consistently lower your grades.

43. How do you typically feel when you have volunteered a comment in class discussion?
    a. You feel very important in the class.
    b. You feel as if you have gained some positive recognition.
    c. You feel neutral about your comment.
    d. You fear that you might have said the wrong thing.
    e. You believe your comment has been poorly received.

44. How much of a personal priority is improving the amount and/or quality of your participation in class discussion?
    a. It is your top priority.
    b. It is among your highest priorities.
    c. You are neutral about the prospect of improving your class participation.
    d. It is among your lowest priorities.
    e. It is a non-priority for you.

45. What is your opinion of the social status of students who participate frequently in class discussion?
    a. They tend to be the most popular students in class.
    b. They are among the more popular students in class.
    c. Frequent participation has little effect on one's standing with peers.
    d. They are among the less popular students in class.
    e. They tend to be the least popular students in class.

46. What is your opinion of the social status of students who participate little, if at all, in class discussion?
    a. They are greatly admired for their quietness.
    b. They are generally admired for their quietness.
    c. Their minimal participation has little effect on how peers regard them.
    d. They are generally discredited for their quietness.
    e. They are greatly discredited for their quietness.

47. What is your view of the relationship between the perceived relevance of course content and student inclination to participate in class discussion?
    a. Students feel the greatest desire to participate in courses they view as highly relevant.
    b. Students feel a moderate desire to participate in courses they view as relevant.
    c. Student inclination to participate is not affected by the perceived relevance of the course content.
    d. Students feel somewhat less need to participate in courses they view as relevant.
    e. Students feel the least need to participate in courses they view as highly relevant.

48. Which of the following best represents how students' participation in class discussion will affect their personal standing with teachers?
    a. Students who participate frequently usually are the most liked by their teachers.
    b. Students who participate frequently increase their chances of being liked by their teachers.
    c. Frequent participation has little effect on how much teachers like a student.
    d. Frequent participation decreases students' chances of being liked by their teachers.
    e. Students who participate frequently are the least liked by their teachers.

49. Which of the following best represents the effect of a teacher's friendliness on student participation in class discussion?
    a. Teacher friendliness is the number one contributor to student participation in class discussion.
    b. Teacher friendliness is among the more important contributors to student participation in class discussion.
    c. Teacher friendliness has little to do with student participation in class discussion.
    d. Teacher friendliness is among the less important contributors to student participation in class discussion.
    e. Teacher friendliness is the least important contributor to student participation in class discussion.

50. Compare the effects of teacher friendliness and teacher knowledge of the subject matter in the course on student participation in class discussion.
    a. Teacher knowledge is a far greater contributor than teacher friendliness to class discussion.

b. Teacher knowledge is a somewhat stronger contributor than teacher friendliness to class discussion.
c. Teacher knowledge and teacher friendliness have an equal impact on class discussion.
d. Teacher friendliness is a somewhat stronger contributor than teacher knowledge to class discussion.
e. Teacher friendliness is a far greater contributor than teacher knowledge to class discussion.

Vita

Lisa Nicole Edge, also published under the name Lisa Foster, was born in Pittsburg, Texas. She was raised in East Texas. She moved to Northeastern Vermont when she was fifteen and attended her final three years of high school in Vermont. She obtained a B.A. in Psychology at Castleton State College in Castleton, Vermont. After graduating, she attended the University of Tennessee's School Psychology Ph.D. program. In December of 2009 Lisa received a M.S. in Applied Educational Psychology. Lisa will receive her Ph.D. in August 2011 after completing a yearlong internship.