Who Benefits from Regular Class Participation?

Lei Tang[1], Shanshan Li[1], Emma Auden[2], Elizabeth Dhuey[3*]

[1]Shaanxi Normal University, 620 West Chang'an Avenue, Xi'an, 710119, China.

[2] Middlebury College, Old Chapel Rd, Middlebury, Vermont, USA.

[3] Department of Management, University of Toronto, 121 St. George Street, Toronto, M5S2E8, Canada.

* Corresponding author. elizabeth.dhuey@utoronto.ca; 416-978-2721.

We would like to thank Prof Aloysius Siow, Prof. Robert McMillan, Prof. Dwayne Benjamin, Prof. Jennifer Murdock, Prof. Robert Gazzale, and Honam Mak for inspiring comments, suggestions, and/or help in experiment conducting and data collection.

Abstract

In this study, we sought to explore the dynamics of in-class participation and its effects on student outcomes. In particular, we looked at three questions: whether students' outcomes were improved by grading participation more intensely; who benefits most from increased participation; and whether the students who would benefit from more intensive grading choose it when they are given the choice. An eight-month field experiment was used to elicit students' preferences for and randomly assign them to different grading intensities. We found that grading students on participation weekly is more effective than biweekly and that students who might be expected to struggle in class (i.e., students who prefer not to be graded weekly, who have lower GPAs or lower self-control scores) benefit most from the weekly participation grading intervention. When students were given a choice, however, the ones who would benefit the most were no more likely to choose weekly grading than were others.

Keywords: active learning, grading intensity, academic achievement, class participation, pedagogy, undergraduate teaching.

Wordcount: 9753

Introduction

Student engagement in active or interactive learning activities is an important predictor of student achievement (Handelsman, Briggs, Sullivan, & Towler, 2005). Various forms of active class participation, such as problem solving periods, class discussion and etc., have been found to be positively correlated with academic performance, critical thinking abilities, and student attitudes toward class participation (Junn, 1994; Garside, 1996; Murray & Lang, 1997; Garard et al., 1998; Handelsman et al., 2005; Rassuli & Manzer, 2005; Wilson, Pollock, &. Hamann, 2007; Lage et al., 2010; Starmer et al., 2015; Godlewska et al., 2019).

Early efforts to increase student engagement with class material most commonly consisted of implementing mandatory attendance policies. The rationale for these types of policies, which tie student grades to attendance, was based on a large literature that found a positive relationship between lecture attendance and academic achievement in higher education (see for example: Van Blerkom, 1992; Gunn, 1993; Romer, 1993; Durden & Ellis, 1995; Devadoss & Foltz. 1996; Marburger, 2001; Dolton et al., 2003; Rocca, 2003; Stanca, 2006; Marburger, 2006; Martins & Walker, 2006; Crede et al., 2010). However, much of this literature suffers from endogeneity issues. A more recent literature uses plausible sources of exogeneous variation and find mixed results. For instance, Krohn and O'Connor (2005); Martins and Walker (2006); Andrietti (2014); and Andrietti and Velasco (2015) find no relationship between attendance and achievement. Arulampalam, Naylor, and Smith (2012) find a positive relationship but only for a subset of students whereas Marburger (2001), Cohn and Johnson (2006), Stanca (2006), Lin and Chen

(2006), Chen and Lin (2008), Dobkin, Gil, and Marion (2010), and Kassarnig et al. (2017) estimate a positive relationship between lecture attendance and academic performance.

More directly related to mandatory class attendance policies, Chan, Shum, & Wright, (1997) and Caviglia-Harris, (2006) conducted randomized experiments using mandatory class attendance policies as a treatment and found little to no effect on grades, despite increasing attendance rates among students. These randomized experimental results indicate the possibility that, even though mandatory attendance policies increase attendance rates, less motivated students may still not pay attention during class and, thus, have no learning improvement, and even when these policies work, they may not work for all students. This explanation is supported by Chen and Lin (2008), who found that students who voluntarily chose to attend class obtained a larger benefit from attending each lecture than did students who skipped some classes.

In more recent years, educators have begun to move beyond policies based simply on attendance to employ new types of grading systems that seek to tie grades to student participation (i.e., participation grading schemes) rather than attendance alone. There are a number of ways that instructors have sought to do this. In some instances, instructors have tried to tie grades to student comments made during class discussions (Lumbantobing, 2012; Nelson, 2010). In other instances, instructors have used Classroom Response Systems (“clickers”), a technology that enables educators to post multiple-choice questions on PowerPoint slides in class and to collect student answers in real time (Barr, 2014).

The goal of participation grading schemes is to encourage students to actively engage with class material and receive more immediate feedback in addition to simply being present in class. Such grading systems aim to both motivate attendance and incentivize active engagement with class materials. Feedback is one of the most powerful influences on learning and achievement (Hattie and Timperly, 2007) and providing active learning activities allows for increased levels of feedback. Bangert-Drowns, et al. (1991) provides a nice meta-analysis of the extensive literature on the effect of feedback in test-like events.

Research on participation grading schemes has generally shown positive outcomes. Freeman et al. (2007) found that in classes in which students answered daily multiple-choice questions by turning in cards or using “clickers”, students had lower failure rates and higher exam scores. Another study found that clickers are especially effective in improving learning outcomes, on average, raising exam scores by one-third of a grade point (Mayer et al., 2009). Research on the effectiveness of clicker-based participation grading schemes has piqued the interest of education scholars, as it provides directions for using technology to increase student engagement with class material (Morling et al., 2008; Mayer et al., 2009; Dawson et al., 2010).

Although studies that demonstrate the positive effects of participation grading schemes make an important contribution to the literature, there are two aspects of participation grading schemes that we believe remain largely unexamined. First, researchers have not yet experimented with how adjusting the intensity of participation may affect the way that students respond to participation grading schemes. Although some studies have compared the

effectiveness of different styles of participation grading, for instance, asking students to answer questions verbally versus using clicker technology (Barr, 2014; Gauci, Dantas, Williams, & Kemm, 2009; Stowell & Nelson, 2007), to our knowledge, no study has examined the effects of adjusting the intensity of a single grading scheme while keeping the style consistent. Varying the intensity of participation demands, for example, requiring students to answer questions during every class versus posting questions only every other week, can help to reveal the mechanisms behind improvements in student performance. Such research also can provide guidelines for educators who hope to implement effective participation grading schemes.

Second, it is unknown whether all students involved in participation grading schemes benefit from them or whether the increase in average learning outcomes can be attributed to improvement by only a subset of students. Although some research has investigated which subgroups of students are most likely be active participants in class and the correlation between voluntary participation and grades (King & Joshi, 2008; Reimer Nili, Nguyen, Warschauer, & Domina, 2015), there has been little rigorous analysis conducted on how student performance may vary by heterogeneous characteristics when students are graded on participation. Identifying whose performance improves when participation grading schemes are enforced can help educators to determine how to adapt grading schemes to fit different student types and learning styles.

Given the absence of research that explores in sufficient depth how participation grading schemes work, we have three objectives in this research. First, we investigate whether adjusting

the intensity of the participation grading scheme (asking clicker questions every week vs. every other week) changes its effectiveness. Second, we consider the heterogeneous effects of our participation grading scheme, observing differences in performance among groups of students with different characteristics. Finally, we examine whether the students who would benefit from more intensive grading choose it when they are given the choice.

To achieve these objectives, we implemented a clicker-based participation grading scheme in a second-year economics course at a large public university in Canada. We designed two different participation grading schemes that vary only in intensity. Students in both schemes are scored on completion and accuracy of questions displayed by PowerPoint during class. However, in one scheme the questions are graded every week (weekly) and in the other scheme it occurred only every other week (biweekly). We randomly assigned students to one of the two grading schemes and compare the results between these two grading schemes to test whether the intensity of participation requirements affects the effectiveness of a participation grading scheme. We then use regression analysis to examine heterogeneous effects, looking for differences in student performance outcomes between students with different preferences about intense (weekly) participation grading, past GPAs, and self-control scores.

After collecting data and running an analysis on a sample of 468 students, we find that our intense (weekly) participation grading scheme led to larger increases in class participation and course grades than did our biweekly grading scheme. The relative effectiveness of intense participation demands on student performance was consistent across all heterogeneous subgroups

in our analysis. However, increasing the intensity of participation demands led to larger improvements among certain subgroups. We find that the effects of the intense weekly participation grading scheme are largest among those students who, during our initial survey, expressed a preference for less intense (biweekly) participation grading. Effects are also large among students who have low prior GPAs and low self-control scores. Our results have important implications in the classroom, as they indicate that students who struggle academically can benefit greatly from intense participation grading schemes.

The contribution of this paper is most clearly seen when compared to findings of other recent studies. Studies such as that of Yourstone, Kraye, and Albaum (2008) and Reimers et al. (2015) show that encouraging students to participate by giving real-time feedback during class periods can improve learning outcomes. Our paper builds on their findings through the implementation of a clicker-based participation grading scheme but goes one step further to compare participation grading schemes of different intensity (weekly vs. biweekly). Our paper also provides a more in-depth analysis of heterogeneous effects. For instance, we investigate student preference for participation grading and then look at the relationship between this and class performance and final learning outcomes. The other heterogeneous variables that we account for—prior GPA and self-control score—are also closely related to the study attitudes and habits of students in our sample. Thus, this is the first paper that tries to identify which subgroups of students are helped most by participation grading based on their study habits and

attitudes. This means that our results are particularly important for educators who hope to adjust their grading schemes to accommodate students with diverse learning backgrounds and needs.

Method

Sampling

The sample comprised students who attended a large public university in Canada during the 2014–2015 academic year. Our main analysis was based on students enrolled in an intermediate quantitative methods class offered within the Faculty of Arts & Science. In total, there were 641 students enrolled in the class, 590 of whom consented to data collection, and 468 of whom did not drop out the class during the semester and were subsequently included in our main analysis. Students were enrolled in one of four sections each term and could switch between the sections for their lecture. Two instructors taught the course, one was responsible for the fall term and the other was responsible for the winter term. Thus, in each term, only one instructor was teaching all four sections.

Randomization

Our randomization process followed a two-step procedure. The first step involved preference elicitation. Before courses began, the students were told that participation grades would be assessed based on student electronic responses to questions posed during class. They were asked to express a *preference* between receiving weekly (once per week) and biweekly (every other week) grades in class participation. They were told that students who chose lottery A

had a higher chance of being assigned to the weekly grading scheme (i.e., there would be more incentive to attend class and to participate frequently). Students who chose lottery B had a higher chance of being assigned to the biweekly grading scheme (i.e., there would be less incentive to attend class and to participate frequently). Students were asked to indicate their preference by choosing between the two lotteries. Because student choices had real consequences in terms of their final assignment, student choices in this lottery should reflect the students' true preferences (desire for more or less incentives to attend class and to participate). Students were also told that their instructors would be blind to their choice of lottery and would know only the final assignment of grading schemes. Among students who consented to data collection, 427 (72%) of the students preferred the weekly grading scheme over the biweekly grading scheme (163 students, 28%).

After preference elicitation, we divided students within each preference group into a treatment or control group. We call the students who were ultimately assigned (after the lottery) to the weekly participation grading group the treatment students. Those who were assigned to the biweekly participation grading group are the control students. When dividing students into treatment and control groups, we randomly assigned 70% of the sample students to their grading scheme of preference and 30% to the scheme that they did not prefer. Figure 1 is a flow chart of the sample selection and randomization, based on students who consented to data collection.

The figure also shows the final sample size of the treatment and control groups. Because group assignments and consent forms were given on the same day, some participants were

dropped after the group assignment because they did not consent to data collection. For this reason, group assignment does not follow exactly the random 70%/30% division that our procedure intended; nevertheless, as discussed below, the final sample remains balanced in all respects.

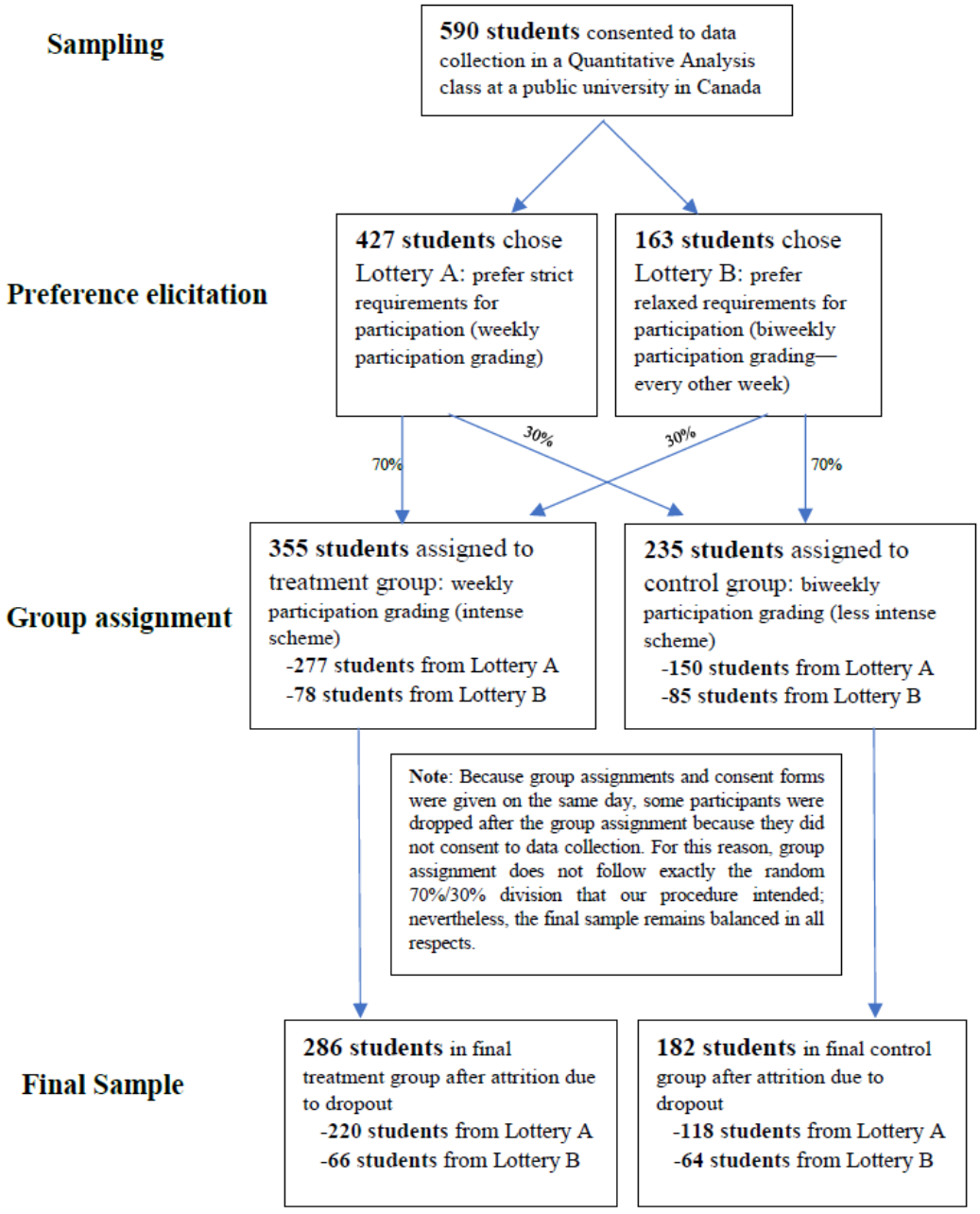

Figure 1. Sampling and group assignment.

Among consenting students, a total of 355 (277 from lottery A, 78 from lottery B) were assigned to the treatment group (the weekly participation grading group), and 235 (150 from lottery A, 85 from lottery B) were assigned to the control group (the biweekly participation grading group). Thus, we essentially have a simple randomized controlled trial (RCT) for each preference group. Among the 590 students who consented to data collection and were assigned to the two grading schemes, 468 students did not drop out of the class and were used for our main analysis.

Intervention

Our intervention involved giving students grades for class participation based on their answers to multiple choice questions asked in class. Although all students were given grades, in part, based on their answers to multiple choice questions, the intensity of the grading scheme differed between the treatment and control groups. Students in the treatment group were graded every week on class participation, while students in the control group were graded biweekly on class participation. The course length was eight months, with 24 regular class meetings, for which each class period lasted for two hours.

The protocol for in-class grading was set in advance and was implemented according to a set schedule that was described on the course syllabus. On average, six questions were asked during the two-hour lecture (with each question being allocated about 3 minutes of time). During graded lectures, students could earn one point for completing each question and an additional

point for answering correctly. Class participation grades counted for 10% of each student's total class grade.

Data Collection

We conducted data collection in three different phases. We first collected data on student attributes. This part of the data collection was conducted through an in-class survey and by collecting administrative data from the university provost office. Next, we carried out the intervention and collected the total score earned by answering questions using their clickers for each student throughout the semester. Finally, we recorded student course grades, which occurred after the course had been completed.

Data on student attributes. We obtained data from administrative records provided by the university that included demographic characteristics of the students in our sample. This information included student gender, year in school, status as a full-time or part-time student, whether English, French or other language was the student's first language, and cumulative grade point average. The data were provided to the research team electronically.

We also conducted an initial survey at the start of the semester. During this survey, we asked students to indicate their preference for either more intense (weekly) or less intense (biweekly) participation grading. The process for collecting student preferences on grading intensity was described above.

The survey was also used to collect data on student personality traits, including measures of self-control, motivation, and risk aversion. First, we measured self-control among the students

in our sample by using two commonly used measures of self-control: the Ideal minus Expected (IE) gap and the International Personality Item Pool (IPIP). The IE gap measure is based on a questionnaire by Ameriks, Caplin, Leahy, and Tyler (2002), in which participants are given hypothetical coupons for free dinners to be used within two years. Participants are asked to state how many coupons they would ideally want to use in the first year (rather than saving them for the second year) and how many they would expect to use in the first year. The IE gap is obtained by subtracting the expected number of coupons used from the ideal number of coupons used; a lower difference represents a larger self-control problem. The other self-control measure, IPIP self-control, is based on an existing psychological inventory, the International Personality Item Pool (IPIP; Goldberg, 1999). It is a single number generated by adding up positive and negative scores from inventory responses. Smaller scores indicate larger self-control problems. Our measure of motivation also is based on the personality questionnaire from the IPIP. The measure is produced as a single number and is generated by summing positive and negative items. The larger the number, the more motivated a student is.

The measure of risk aversion was based on a scale developed by Holt and Laury (2002). We used two measures for risk aversion. The first asked students to make 11 decisions between risky and safe options. The higher the number of safe choices, the more risk averse a student is. The second measure asked students to rate their general attitude toward risk-taking, asking, “In general, on a scale from 0 to 10, how willing are you to take risks? 0 means unwilling to take risks, and 10 means being fully prepared to take risks.”

All measures of student personality traits (self-control, motivation, and risk aversion) were standardized by class means and class standard errors for the purposes of our main analysis. We supplemented that information with measures of student difficulties in attending class. This was based on student survey responses about travel distance to school as well as work, family, and social commitments.

Participation scores. We collected participation grades for each student throughout the semester. Participation was graded using clickers to electronically record and grade student responses to class questions. Students earned one point for completing each question and an additional point if they answered correctly. Two variables were generated from the participation scores. One is class participation rate, which is a ratio of total numbers of classes that a student participated in at least one question during a class to the total number of classes. This aims to measure students' efforts to attend class. Another is clicker score, which is an average of participation scores over the whole course. Clicker score is used as a proxy for student effort in answering questions in classes. Students needed to have the hand-held clicker device to participate in class, as clicker availability could affect rates of compliance. Due to this requirement, students needed to purchase or borrow and use the hand-held device to answer questions in the class. Students had three options: (a) purchase a new clicker and sell it back to the university bookstore at the end of the course, (b) purchase a used clicker, or (c) borrow a clicker from the department or from the instructor before class. There were no students who did not participate due to not having a clicker (as the instructor always had enough spares in class).

Course grades. Finally, we collected data on overall course grades for each student. These data were obtained from the course instructors. These data, along with the participation rate and clicker scores, became the dependent variables used in our analysis.

Balance and Attrition

Table 1 shows that 122 students (21% of the 590 consenting students) dropped out the class, but these dropouts were not statistically significantly different from the 468 non-dropouts (79% of the 590 consenting students) in terms of pre-treatment characteristics.

Table 1. Balance Check between Dropouts and Non-Dropouts

Dropout status	Obs.	Prefer weekly grading (%)	Assigned to weekly grading (%)	CGPA	Year of study	Female	Self-control (IE gap)	Motivation (IPIP)	Risk aversion (*n* of safe choices)
Non-dropout	468	72%	61%	3.11	2.11	59%	0.04	16.61	5.3
Dropout	122	73%	57%	2.71	2.06	55%	0.04	18.23	5.46
p-value				0.0000	0.2093	0.4298	0.9886	0.0077	0.4640
Obs.	590	590	590	590	585	589	590	590	547

Students assigned to the weekly grading scheme (treatment group) were approximately 8% less likely to drop out than were those assigned to the biweekly grading scheme (control group), a difference that is statistically significant ($p < .05$). However, as shown in Panels A and B of Table 2, among the 468 non-dropouts, the treatment and control groups were not significantly different in terms of the pre-treatment characteristics. Thus, the baseline characteristics of our treatment and control groups are balanced both before and after attrition.

Table 2. Balance Check between Treatment and Control Groups by Preference Types

Panel A: Students who Prefer Biweekly Grading

Assignment of grading scheme	%	CGPA	Failed course(s) before	Year of study	Female	Self-control (IE gap)	Motivation (IPIP)	Risk Aversion (*n* of safe choices)
Weekly	49%	3.12	17%	2.1	49%	-0.52	16.3	5.15
Biweekly	51%	3.08	19%	2.1	56%	-0.07	17.22	5.3
p-value		0.7443	0.6634	0.8331	0,4286	0.2592	0.4345	0.7049
Obs.	130	130	127	129	129	123	130	119

Panel B: Students who Prefer Weekly Grading

Assignment of grading scheme	%	CGPA	Failed course(s) before	Year of study	Female	Self-control (IE gap)	Motivation (IPIP)	Risk Aversion (*n* of safe choices)
Weekly	65%	3.1	15%	2.1	60%	0.13	16.31	5.4
Biweekly	35%	3.13	16%	2.1	64%	0.27	17.02	5.2
p-value		0.7622	0.8477	0.9481	0.5234	0.5445	0.2697	0.3771
Obs.	338	338	326	338	338	323	338	321

Source: Authors' tabulation of 2014 registrar office and survey data.

Statistical Analysis

Treatment effects. Due to the experimental design, a higher proportion of those who preferred the weekly grading scheme were actually assigned to it than those who did not prefer it. This means that the treatment status is not random in the overall sample even it is random within each of the two preference groups. If the treatment effects are different across the preference groups, we can arrive at a biased estimate of the average treatment effect (ATE). To identify ATEs of the participatory grading scheme, we calculate the heterogeneous treatment effects for the two preference groups by running a regression with dummies for all the strata (i.e., whether a student preferred the weekly scheme) and interaction terms with the treatment

dummy . We then calculate the ATE by creating a weighted average of the different treatment effects within each preference group. This will get us the same heterogeneous effects for each preference type and the ATEs as we do simple regressions for each preference group separately and then create a weighted average for the ATE (see Athey and Imbens, 2017). The regression specification is as follows:

$$Perform_i = \alpha + \beta_1 Pref_i + \beta_2 Pref_i \times Weekly_i + \beta_3 NotPref_i \times Weekly_i + X_i\hat{\gamma} + \epsilon_i \quad (1)$$

The dependent variable, $Perform_i$, is one of our three performance measures: class participation rates, clicker scores (a proxy for effort in answering in-class questions), and overall course grades. $Pref_i$ is an indicator of whether the student preferred the weekly grading scheme, $NotPref_i$ is an indicator of whether the student preferred the biweekly grading scheme, $Weekly_i$ is an indicator of whether the student was assigned to the weekly grading scheme. $Pref_i \times Weekly_i$ is an interaction term, which equals one for those who preferred and assigned to the weekly grading scheme, and $NotPref_i \times Weekly_i$ is an interaction term, which equals one for those who preferred the biweekly grading and were assigned to the weekly scheme.

The coefficients in our regression thus represent the different possibilities for student preference for and assignment to the two grading schemes. The constant α measures the performance of students who both preferred and were assigned to the biweekly scheme; β_1 measures the performance of students who preferred weekly grading but were assigned to biweekly grading, β_2 measures the treatment effects for students who preferred and were

assigned to the weekly scheme, and β_3 measures treatment effects for students who preferred biweekly grading but were assigned to the weekly scheme. We also conducted a *t*-test for the equality of coefficients, $\beta_2 = \beta_3$. This test allows us to estimate whether being assigned to the weekly grading scheme has different effects on the performance of students with different preferences for grading intensity. To increase the efficiency of the analysis, we also controlled for a set of pre-treatment student attributes, represented by the vector X_i, in the regression. These attributes include measures of student academic achievement before this study (e.g., cumulative GPA), student demographic information (e.g., gender, year of study, full-time study indicator, first language, program of study), as well as survey measures of self-control, motivation, risk aversion, and personal difficulties in attending classes.

Based on the coefficient estimates of Equation (1), the overall treatment effects are weighted averages calculated from averaging the treatment effect of students in both preference groups. Specifically, we generate the overall treatment effects by weighting the treatment effects of those who preferred the weekly grading (β_2) and the treatment effects of those who preferred the biweekly grading (β_3) by the proportion of students in each of the two preference groups.

We also conducted heterogeneous treatment effects on students with different prior academic achievement, measured by student GPA levels before the experiment. We divided students into three GPA categories: students with prior GPAs in the highest range (3.5 and above, denoted as $HGPA_i$); students in the middle range (between 2.5 and 3.4, denoted as $MGPA_i$); and students in the lowest range (2.4 and below, denoted as $LGPA_i$). In the regression, students

in the middle range are omitted as the reference group. Due to the experimental design, a higher proportion of those who preferred the weekly grading scheme were actually assigned to it than those who did not prefer it. This means that the treatment status is not random in the overall sample even it is random within each of the two preference groups. Thus, we calculate the ATE by creating a weighted average of the different treatment effects within each preference group similarly as in regression (1). The full regression model is as follows:

$$Perform_i = \rho_0 + \rho_1 Pref_i + \rho_2 Weekly_i + \rho_3 LGPA_i + \rho_4 HGPA_i + \rho_5 Pref_i \times Weekly_i + \rho_6 LGPA_i \times Weekly_i + \rho_7 HGPA_i \times Weekly_i + \rho_8 LGPA_i \times Weekly_i \times Pref_i + \rho_8 HGPA_i \times Weekly_i \times Pref_i + X_i\hat{\gamma} + \epsilon_{pi} \quad (2)$$

Similarly, we conducted a heterogeneous treatment effect on students with different self-control abilities. The specifications of the other analyses that examine heterogeneous treatment effects are similar to Equation (2).

Who preferred the weekly grading scheme? To understand more about the mechanism behind the effects of participation grading, we need to understand what types of students are affected by more intense participation grading. To do this, we conducted a probit regression.

$$\Phi^{-1}(Pref_i) = \delta_0 + X_i\hat{\delta} + \epsilon_i \quad (3)$$

The dependent variable, $Pref_i$, is a dummy variable that represents student i's preference for a weekly grading scheme, in which 1 indicates a preference for weekly grading and 0 represents a preference for biweekly grading. X_i is the same vector of explanatory variables (the same as in Equation (1). We examine the two sets of characteristics. One set are observable characteristics,

specifically, student cumulative GPAs before this study and student demographic information (e.g., age, gender, year of study, parental education, first language, program of study). The other set are usually not directly observable characteristics, specifically, measures of self-control, motivation, risk aversion, and personal difficulties in attending classes. The vector of the estimated coefficient, $\hat{\delta}$, tells us whether there is a relationship between preference for grading intensity and student characteristics.

Results

The Overall Treatment Effect

Our analysis found that, overall, treatment students who experienced the more intense, weekly participation grading scheme had significantly higher participation rates and educational performance than did control group students who were graded every other week on participation (Table 3). Specifically, the treatment of weekly grading (compared to biweekly grading) led to a 10.96% increase in class participation rates and a 2.75% increase in clicker scores. The treatment also increased the overall course grades, leading to an increase of 6.31%.

Table 3. Overall Treatment Effects

Variable	Class participation rate	Clicker score	Course grade
Treatment effect	10.96**	2.75**	6.31**
	(1.868)	(0.708)	(1.596)
Control group performance	73.47**	73.65**	64.33**
	(1.597)	(0.594)	(1.413)
Observations	468	467	468

Note. Outcome variables are scores out of 100; robust standard errors are in parentheses; overall treatment effects are weighted averages calculated from averaging the treatment effect of students in both preference groups. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Heterogeneous Effects

Although Table 3 shows a significant average effect from the treatment, these results do not fully capture how the treatment may have affected different subgroups of students. To show this, we initially divided the sample by preference group and compared the treatment effects. We also separated the sample by prior GPA and self-control scores to investigate the heterogeneous effects of treatment on these different subgroups.

Treatment Effects by Lottery Preference

Table 4 provides a summary of the regression results of Equation (1). We separated students into "prefer weekly grading," which refers to those students who selected lottery A and

Table 4. Treatment Effect by Preference Type

Preference	Class participation rate	Clicker score	Course grade
Prefer weekly grading			
Control group performance	78.05**	75.01**	68.27**
	(1.866)	(0.657)	(1.750)
Treatment effects	7.09***	1.62**	2.58
	(2.140)	(0.788)	(1.970)
Prefer biweekly grading			
Control group performance	63.28**	70.62**	55.55**
	(3.051)	(1.237)	(2.364)
Treatment effects	19.60**	5.25**	14.60**
	(3.706)	(1.465)	(2.700)
Observations	468	467	468
Adjusted R^2	0.13	0.065	0.091
RMSE	18.40	7.16	15.51

Note. Outcome variables are scores out of 100; robust standard errors are in parentheses.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

who opted for a more intense grading system, and “prefer biweekly grading,” which refers to those students who preferred to have a less intense participation grading scheme.

The results in Table 4 show that the treatment effect (being in the weekly grading system) is much larger and more significant for the group that preferred biweekly grading. In other words, those who preferred *not* to have intense participation grading actually benefited the *most* from it. After dividing the sample by preference group, we still see a positive treatment effect in both groups in class participation rate and clicker score. The treatment effect, however, is markedly larger among the group that preferred biweekly grading. In the group that preferred biweekly grading, being in the treatment group (weekly grading) led to an increase of 19.60% and 5.24% in class participation rates and clicker scores, respectively, vs. 7.09% and 1.62% in the group that originally preferred weekly grading. Treatment also led to a significant 14.60% improvement in course grades among those students who preferred biweekly grading. The increase in course grades among students who preferred weekly grading, however, was not significant.

Treatment effects by prior GPA. To examine how the effects of the weekly participation grading scheme differ based on prior academic performance, we performed a heterogeneous analysis based on prior GPA of students in our sample. Table 5 shows the results for the treatment effect between sample students with prior GPAs in the highest range (3.5 and above) and students in the lowest range (2.4 and below).

Table 5. Treatment Effect by Prior GPA

Prior GPA	Class participation rate	Clicker score	Course grade
Highest GPA			
Control group performance	85.36**	77.60**	81.24**
	(2.105)	(0.805)	(1.544)
Treatment group performance	85.78**	78.29**	76.73**
	(1.808)	(0.806)	(1.355)
Treatment effect	0.41	0.69	-4.51**
	(2.775)	(1.139)	(2.054)
Lowest GPA			
Control group performance	55.70**	68.93**	46.11**
	(3.769)	(1.377)	(2.246)
Treatment group performance	75.47**	74.07**	64.58**
	(2.548)	(0.965)	(1.393)
Treatment effect	19.78**	5.14**	18.47**
	(4.550)	(1.682)	(2.643)
Observations	468	467	468
Adjusted R^2	0.242	0.136	0.348
RMSE	17.18	6.879	13.14

Note. Outcome variables are scores out of 100; robust standard errors are in parentheses; results are weighted averages calculated from averaging the treatment effect of students in both preference groups.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Table 5 shows that treatment effect was much larger among students who had a low prior GPA. For low-GPA students, treatment led to a rise of 19.78% in class participation rates and a 5.14% increase in clicker scores. In contrast, there was no significant treatment effect on participation rates or clicker scores among students with a high prior GPA. In terms of course grades, there was actually a slightly significant negative treatment effect of -4.51% for students with a high prior GPA. For students with a low prior GPA, however, the treatment led to an increase of 18.47% in course grades. The control group's average course grade was 46.11%,

whereas, for the treatment group, it was 64.58%. Thus, for students with low prior GPAs, the treatment effectively brought their scores out of the failing range (on average), as 50% was the cutoff score for failing the class.

Treatment effects by self-control abilities. We also found that students who scored lower on the self-control indices (less self-control) benefited more from the intense weekly participation grading scheme. The treatment effect on class participation rates, clicker scores, and course grades were all significant among both students with the highest self-control scores and students with the lowest self-control scores. Treatment effects for students with high self-control, however, were lower (10.48%, 3.33%, and 6.09% increases in participation rates, clicker scores, and course grades, respectively) in comparison to treatment effects for students with low self-control (14.26%, 4.13%, and 9.67%, respectively). As seen in Table 6, it is clear that, although students with lower self-control perform worse than do those with higher self-control when they are not assigned to the weekly grading scheme, their performances can be increased to a level comparable to that of students with the most self-control if they are assigned to the weekly grading.

Despite the fact that students with lower self-control benefited more from the weekly participation grading scheme, they were actually less likely to prefer it. Our data show that students whose self-control was one standard deviation higher than the class average were about 4% more likely to prefer the weekly grading than were average students.

Table 6. Treatment Effect by Self-Control Score

Self-control	Class participation rate	Clicker score	Course grade
Highest self-control			
Control group performance	71.68***	72.97***	61.89***
	(3.098)	(31.110)	(2.438)
Treatment group performance	82.16***	76.30***	67.98***
	(2.370)	(0.876)	(2.018)
Treatment effect	10.48***	3.33***	6.09**
	(3.047)	(31.109)	(2.543)
Lowest self-control			
Control group performance	67.38***	71.41***	56.45***
	(3.721)	(1.229)	(13.113)
Treatment group performance	81.65***	75.54***	66.11***
	(2.496)	(0.978)	(1.824)
Treatment effect	14.26***	4.13***	9.67***
	(3.802)	(1.350)	(3.177)
Observations	468	467	468
Adjusted R^2	0.124	0.071	0.114
RMSE	18.46	7.136	15.31

Note. Outcome variables are scores out of 100; robust standard errors are in parentheses; results are weighted averages calculated from averaging the treatment effect of students in both preference groups.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Who Preferred the Weekly Grading Scheme

The probit regression results of Equation (3), reported here, show whether there are correlations between student preferences for weekly grading intensity and individual attributes. Table 7 Panel A provides those correlations. The results show that female students and students whose first language was not English or French were more likely to prefer weekly participation

grading. There was no statistically significant correlation between preference for weekly grading and prior GPA.

Table 7 shows, however, that there was a statistically significant correlation between student preferences for grading intensity and their scores on the self-control ability scale. Specifically, students whose self-control was one standard deviation higher than the class average were about 4% more likely to prefer the weekly grading than were average students. This indicates that, even though students with lower levels of prior achievement and less self-control would perform better if they were assigned to the weekly grading scheme, they did not choose the more helpful option when they were given the choice. These results are robust when we use OLS by imposing the linearity assumption and probit regression by relaxing the linearity assumption.

Table 7. Correlation between Preference for Weekly Grading and Student Characteristics

Panel A: Preference and observable characteristics	
Variable	Prefer weekly
Cumulative GPA	-0.003
	(0.021)
Female	0.075*
	(0.043)
Year of study	-0.019
	(0.051)
Full-time indicator	0.069
	(0.077)
First Language not English or French	-0.078*
	(0.044)
Observations	444

Panel B: Preference and unobservable characteristics

Variable	Prefer weekly
Self-control (IE gap)	0.041*
	(0.022)
Motivation	0.022
	(0.027)
Risk aversion (*n* of safe choices)	0.013
	(0.022)
Major obstacles to attending class	
(Omitted category: no major obstacles)	
Work commitment	-0.169**
	(0.071)
Family obligation	0.026
	(0.088)
Travel distance	0.094
	(0.058)
Social commitment	0.078
	(0.066)
Observations	439

Note. Robust standard errors in parentheses; Cumulative GPA, Self-control, Motivation, and Risk Aversion are standardized by class means and standard errors; Work, family, travel, and social obstacles are dummy variables. Values are marginal probabilities measured at means.
† $p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Robustness Check

In this section, we present the results of a robustness check for the treatment effects when we use alternative measures of self-control and prior academic achievement. Table A1 in the appendix shows the heterogeneous effects by self-control when we use alternative measures of self-control, IPIP self-control. The result is similar to the results when IE gap was used as the measure of self-control, which means that our conclusion regarding self-control is robust.

Table A2 in the appendix shows the heterogeneous effects by prior academic achievement when we measure academic achievement by a dummy variable of whether a student failed the course before. The results show that weekly grading improved class participation for students who failed or did not fail courses before but improved course grades of only those who did not fail courses before (i.e., only better students were helped by the weekly grading). Note that this is inconsistent with the results when cumulative GPA was used as the measure of academic achievement. This result, however, based on the indicator of whether a student failed courses before, may be unreliable, as the proportion of students who failed courses before is much smaller than of those who did not fail before.

This study can also provide possible channels through which the more intense participation grading could increase student performance. Did the more intense participation grading increase learning through its effect on class attendance rather than other student effort, such as self-study times? This study found that the more intense participation grading did not significantly increase self-study hours (the difference in weekly self-study hour = - 0.16, SD = 0.185). This result indicates the possibility that the weekly grading scheme increased learning through increased class participation rates and increased effort in answering questions in class rather than through its effects on self-study efforts. However, this is not conclusive since the self-study hour measures were measured at the early part of the course and were self-reported by students.

Discussion and Conclusion

Implications of the Findings

Our research shows the results of an RCT in which two participation grading schemes were implemented: one intense weekly participation grading scheme, which we refer to as the treatment, and one less-intense biweekly participation grading scheme, which we refer to as the control. Our results show that the treatment had a positive effect on student participation rates and course grades. After conducting a heterogeneous analysis, we found that treatment was especially effective among students who preferred biweekly (less intense) participation grading, who had lower prior GPAs, and who had less self-control.

The results of our study are consistent with past research that shows that participation grading can effectively raise participation levels and raise student academic performance. Researchers have found that grading student participation through multiple-choice questions administered during class can lower failure rates and improve exam scores (Freeman et al., 2007). Several studies also have emphasized the effectiveness of using clicker technology to ask questions in class, showing that it leads to significant improvement in exam scores and grades (Mayer et al., 2009; Yourstone et al., 2008; Reimer et al., 2015). Our study contributes to this literature by confirming that clicker technology can be used effectively to implement participation grading.

Our finding is also consistent with the research that shows more frequent classroom testing increases students' exam performance (Bangert-Drowns et al., 1991). This study is

different from these early studies by investigating how changing the intensity of an in-class participation grading scheme instead of the intensity of classroom tests can affect student learning. Our results show that grading students weekly on participation is significantly more effective than grading students every other week on participation. This demonstrates that frequent implementation of participation grading is essential to its effectiveness. The study design has only two levels of intensity (weekly and bi-weekly) and thus do not provide an answer to the level of optimal intensity. Further research could explore more variation in the intensity of participation grading to determine the most efficient way to implement such a grading scheme.

The effectiveness of more intense participation grading could be due to a variety of reasons cited in the clicker-related literature. Researchers have found that clickers are effective in improving student learning, in large part, because they promote student interactions with peers and teachers about class material (Blasco-Arcas, Buil, Hernández-Ortega, & Sese, 2013). Thus, the effectiveness of intense participation grading in our study may be due to more frequent discussions of class material between peers and instructors or a more conducive learning environment formed by frequent interactions. In a similar vein, scholars have noted that clickers provide teachers with a means by which to assess student understanding of concepts in real time during class (d'Inverno, Davis, & White, 2003; Roschelle, Penuel, & Abrahamson, 2004; Caldwell, 2007). Teachers may tailor their teaching to student needs because they have a better gauge of student understanding; for example, an instructor may choose to provide further

explanation or move on from a concept, depending on class clicker responses. Alternatively, student retention of class material may have been higher during lectures with clicker questions. Clicker questions provide variety in a lecture and may serve as a break that allows students to refocus their attention, thus leading to better learning outcomes (Middledorf & Kalish, 1996).

Our study is also the first study to look at how student attitudes toward participation grade schemes change the intervention's effectiveness. Although past research has collected self-reported data on whether students liked clickers or found them helpful (DeBourgh, 2008; Draper & Brown, 2004; Johnson & Lillis, 2010; Patterson, Kilpatrick, & Woebkenberg, 2010), scholars have not examined attitudes toward clickers or participation grading in relation to their impact on student performance. Our study shows that students who prefer less intense participation grading actually benefit the most from intense participation grading. Students in this group may need structured, external motivation to succeed academically but are not aware of this or are not willing to sign up for the extra work. Educators should be aware of this when designing incentives for students, as students may not have the knowledge or motivation to actively pursue the style of learning that benefits them the most.

The heterogeneous effects analysis in this paper shows that students with lower prior GPAs and lower self-control scores benefit disproportionately from the enforcement of weekly participation grading during class. This conclusion about self-control is broadly consistent with previous studies that show that self-imposed constraints at school or work increased task performance among people who lacked self-control (Ariely & Wertenbroch, 2002; Webb,

Christian, & Armitage, 2007; Kaur, Kremer, & Mullainathan, 2014). Our heterogeneous results are also consistent with those of Freeman et al. (2007), who found that students at high risk of dropping out of an introductory biology class benefited disproportionately from the implementation of a participation grading scheme. It seems that students who are most disadvantaged when coming into a class tend to benefit the most from the imposition of strict requirements for class participation.

The results of our heterogeneous analysis are especially important, given that previous research suggests that the most motivated students benefit most from mandatory attendance policies (Chen & Lin, 2008). Mandatory attendance policies may force students to come to class but do little to encourage struggling students to pay attention or engage with class material. In contrast, the participation grading scheme used in our study is especially effective in improving scores among students who have struggled previously in school.

Limitations and Conclusion

As our treatment was randomized within the same class, and students could attend lectures of any section of the class, student interaction might be a concern (Chen & Lin, 2015). Because more students were assigned to the weekly grading, the class-participation decisions of students who were not assigned to weekly grading might be influenced by those who were assigned to it. The treatment effects, however, would be very likely biased downward, as those who were assigned to biweekly grading might come to more classes or participate more if many of their friends or peers were required to do so. In addition, the study does not take into account

whether study efforts in other classes were affected. If the participation incentives increase performance only in the treated class, but crowd out study efforts in other classes, then the intervention will not help disadvantaged students to improve their overall academic performance or to graduate on time.

Due to data limitation, the study cannot disentangle whether the positive treatment effect comes from an improvement of the critical thinking process (reflected in the correctness component of clicker scores) or from the encouragement to show up in class and participate by answering the questions, regardless of correctness. The study also cannot answer questions of whether the treatment effect is driven by teacher-student interaction (such as the instant feedback that students got right after they answered a question) or by the student-student interaction (such as student discussions with neighboring students before they submit own answers). Further studies are needed to explore the possible channels driving the treatment effects.

We find that grading students on participation once a week is more effective than grading participation every other week. In addition, our heterogeneous analysis shows that students who prefer not to be graded weekly, who have lower previous GPAs or lower self-control scores benefit most from the weekly participation grading intervention. The estimated effects in our study are short-term and are specific to university students enrolled in a single course. To further investigate the longer-term effects and external validity of this study in other disciplines or learning environments, future research would benefit from including more longitudinal studies across different disciplines and different teaching styles.

References

Ameriks, J., Caplin, A., Leahy, J., & Tyler, T. (2002). Measuring self-control problems. *The American Economic Review*, *97*(3), 966–972.

Andrietti, V. (2014). Does lecture attendance affect academic performance? Panel data evidence for introductory macroeconomics. *International Review of Economics Education*, 15, 1-16.

Andrietti, V. & Velasco, C. (2015). Lecture attendance, study time, and academic performance: a panel data study. *The Journal of Economic Education*, 46(3), 239-259.

Ariely, D., & Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science*, *13*(3), 219–224.

Arulampalam, W., Naylor, R., & Smith, J. (2012). Am I missing something? The effects of absence from class on student performance. *Economics of Education Review*, 31(4), 363-375.

Athey, S. and Imbens, G.W. (2017). The Econometrics of Randomized Experiments. In *Handbook of Economic Field Experiments*. 1, ed. A.V. Banerjee, E. Duflo, 73-140. Amsterdam: North-Holland.

Bangert-Drowns, R. L., Kulik, J. A., Kulik, C. L. C. (1991). Effects of frequent classroom testing. *The Journal of Educational Research*, 85(2): 89-99.

Barr, M. L. (2014). Encouraging college student active engagement in learning: The influence of response methods. *Innovative Higher Education*, *39*, 307–319.

Blasco-Arcas, L., Buil, I., Hernández-Ortega, B., & Sese, F. J. (2013). Using clickers in class: The role of interactivity, active collaborative learning and engagement in learning performance. *Computers & Education, 62*, 102–110.

Caldwell, J. E. (2007). Clickers in the large classroom: Current research and best-practice tips. *CBE—Life Sciences Education, 6*(1), 9–20.

Caviglia-Harris, J. L. (2006). Attendance and achievement in economics: Investigating the impact of attendance policies and absentee rates on student performance. *Journal of Economics and Finance Education, 4*(2), 1–15.

Chan, K. C., Shum, C., & Wright, D. J. (1997). Class attendance and student performance in principles of finance. *Financial Practice and Education, 7*, 58–65.

Chen, J. & Lin, T. F. (2008). Class attendance and exam performance: A randomized experiment. *Journal of Economic Education, 39*(3), 213–227.

Chen, J. & Lin, T-F. (2015). Effect of Peer Attendance on College Students' Learning Outcomes in a Microeconomic Course. *The Journal of Economic Education,* 46(4), 350-359.

Cohn, E. & Johnson, E. (2006). Class attendance and performance in principles of economics. *Education Economics,* 14(2), 211-233.

Crede, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research, 80*(2), 272–295.

Dawson, D.L., K.N. Meadows, & T. Haffie. 2010. The Effect of Performance Feedback on Student Help-Seeking and Learning Strategy Use: Do Clickers Make a Difference? The Canadian Journal for the Scholarship of Teaching and Learning. 1 (1): 1-20.

DeBourgh, G. A. (2008). Use of classroom “clickers” to promote acquisition of advanced reasoning skills. *Nurse Education in Practice*, *8*(2), 76–87.

Devadoss, S., & Foltz, J. (1996). Evaluation of factors influencing student class attendance and performance. *American Journal of Agriculture Economics*, *78*(3), 499–507.

D’Inverno, R., Davis, H., & White, S. (2003). Using a personal response system for promoting student interaction. *Teaching Mathematics and Its Applications: International Journal of the IMA*, *22*(4), 163–169.

Dobkin, C., Gil, R., & Marion, J. (2010). Skipping class in college and exam performance: evidence from a regression discontinuity classroom experiment. *Economics of Education Review*, 29, 566-575.

Dolton, P., Marcenaro, O., & Navarro, L. (2003). The effective use of student time: A stochastic frontier production function case study. *Economics of Education Review*, *22*(6), 547–560.

Draper, S. W., & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted Learning*, *20*(2), 81–94.

Durden, E., & Ellis, L. (1995). The effects of attendance on student learning in principles of economics. *American Economic Review*, *85*(2), 343–346.

Freeman, S., O'Connor, E., Parks, J. W., Cunningham, M., Hurley, D., Haak, D., . . . Wenderoth, M. P. (2007). Prescribed active learning increases performance in introductory biology. *CBE—Life Sciences Education*, *6*(2), 132–139.

Garard, D. L., Hunt, S. K., Lippert, L., & Paynton, S. T. (1998). Alternatives to traditional instruction: Using games and simulations to increase student learning and motivation. *Communication Research Reports*, *15*, 36–44.

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.

Gauci, S. A., Dantas, A. M., Williams, D.A. Williams, & Kemm, R. E. (2009). Promoting student-centered active learning in lectures with a personal response system. *Advances in Physiology Education*, *33*, 60–71.

Godlewska, A., Beyer, W., Whetstone S., Schaefli, L., Rose J., Talan, B., Kamin-Patterson, S., Lamb, C., & Forcoine, M. (2019). Converting a large lecture class to an active blended learning class: why, how, and what we learned. *Journal of Geography in Higher Education*, 43(1), 96-115.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe*, 7, 7-28. Tilburg, The Netherlands: Tilburg University Press.

Gunn, P. (1993). A correlation between attendance and grades in a first-year psychology course. *Canadian Psychology*, *34*(2), 201–202.

Handelsman, M. M., Briggs, W. L., Sullivan, N., & Towler, A. (2005). A measure of college student course engagement. *The Journal of Educational Research*, *98*, 184–191.

Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81-112.

Holt, C.A. and Laury, S.K. (2002). Risk Aversion and Incentive Effects. *The American Economics Review*, 92(5), 1644-1655.

Johnson, K., & Lillis, C. (2010). Clickers in the laboratory: Student thoughts and views. Interdisciplinary *Journal of Information, Knowledge, and Management*, *5*(1), 139–151.

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

Kassarnig, V., Bjerre-Nielsen, A., Mones, E., Lehmann, S., Lassen, D.D. Class attendance, peer similarity, and academic performance in a large field study. *PLoS ONE*, 12(11), e0187078. https://doi.org/10.1371/journal.pone.0187078

Kaur, S., Kremer, M., & Mullainathan, S. (2014). Self-control at work. *Journal of Political Economy*, *123*(6), 1227–1277.

King, D.B. & Joshi, S. (2008). Gender differences in the use and effectiveness of personal response device. *Journal of Science Education and Technology*, 17(6), 544-552.

Krohn, G.A. & O'Connor, C.M. (2005). Student effort and performance over the semester. *The Journal of Economic Education*, 36(1), 3-28.

Lage, M.J., Platt, G.J. & Treglia, M. (2010). Inverting the classroom: a gateway to creating an inclusive learning environment. *The Journal of Economic Education*, 31(1), 30-43.

Lin, T.-F., & Chen, J. (2006). Cumulative class attendance and exam performance. *Applied Economics Letters*, 13(14), 937-942.

Lumbantobing, R. (2012). All in: using poker chips to maximize class participation and facilitate discussion. Available at SSRN: http://ssrn.com/abstract=2102045 or http://dx.doi.org/10.2139/ssrn.2102045

Marburger, D. R. (2001). Absenteeism and undergraduate exam performance. *Journal of Economic Education*, *32*, 99–110.

Marburger, D.R. (2006). Does mandatory attendance improve student performance? *Journal of Economic Education*, 37(2), 148-155.

Martins, P., & Walker, I. (2006). Student achievement and university classes: Effects of attendance, size, peers, and teachers. (*IZA Discussion Paper 2490*, Institute for the Study of Labor).

Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., Zhang, H. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology*, *34*(1), 51–57.

Middendorf, J., & Kalish, A. (1996). The “change-up” in lectures. *The National Teaching and Learning Forum*, 5(2), 1–5.

Morling, B., M. McAuliffe, L. Cohen, T. M. DiLorenzo. (2008). Efficacy of Personal Response Systems (‘Clickers’) in Large, Introductory Psychology Classes. Teaching of Psychology. 35 (1), 45-50

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

Nelson, K. G. (2010). Exploration of classroom participation in the presence of a token economy. *Journal of Instructional Psychology*, *37*(1), 49–56.

Patterson, B., Kilpatrick, J., & Woebkenberg, E. (2010). Evidence for teaching practice: The impact of clickers in a large classroom environment. *Nurse Education Today*, *30*(7), 603–607.

Rassuli, A. (2011). Engagement in classroom learning: Creating temporal participation incentives for extrinsically motivated students through bonus credits. *Journal of Education for Business*, *87*(2), 86–93.

Reimer, L. C., Nili, A., Nguyen, T., Warschauer, M., & Domina, T. (2016). Clickers in the wild: A campus-wide study of student response systems. *Transforming Institutions: 21st Century Undergraduate STEM Education*, 383.

Rocca, K. (2003). Student attendance: A comprehensive literature review. *Journal on Excellence in College Teaching*, *14*(1), 85–107.

Romer, D. (1993). Do students go to class? Should they? *The Journal of Economic Perspectives*, 7(3), 167-174.

Roschelle, J., Penuel, W. R., & Abrahamson, L. (2004). The networked classroom. *Educational Leadership*, *61*(5), 50–54.

Stanca, L. (2006). The effects of attendance on academic performance: Panel data evidence for introductory microeconomics. *The Journal of Economic Education*, *37*(3), 251–266.

Starmer, D .J., Duquette, S., Howard, L. (2015). participation strategies and student performance: An undergraduate health science retrospective study. *The Journal of Chiropractic Education*, *29*(2), 134–138. doi:10.7899/JCE-14-20.

Stowell, J. R., & Nelson, J. M. (2007). Benefits of electronical audience response systems on student participation, learning, and emotion. *Teaching of Psychology*, *34*(4), 253–258.

Van Blerkom, L. (1992). Class attendance in an undergraduate course. *Journal of Psychology*, *126*(5), 487–494.

Webb, T. L., Christian, J., & Armitage, C. J. (2007). Helping students turn up for class: Does personality moderate the effectiveness of an implementation intention intervention? *Learning and Individual Differences*, *17*, 316–327.

Wilson, B. M., Pollock, P. H., & Hamann, K. (2007). Does active learning enhance learner outcomes? Evidence from discussion participation in online classes. *Journal of Political Science Education*, *3*(2), 131–142.

Yourstone, S. A., Kraye, H. S., & Albaum, G. (2008). Classroom questioning with immediate electronic response: Do clickers improve learning? *Decision Sciences Journal of Innovative Education*, *6*(1), 75–88.

Appendix

Table A1. Heterogeneous Treatment Effects by IPIP Self-Control

	Full sample		
Variable	Class participation rate	iClicker score	Course grade
Highest self-control			
Control group performance	81.08**	75.98**	68.92**
	(3.363)	(1.292)	(3.004)
Treatment group performance	81.61**	74.17**	67.90**
	(3.139)	(1.374)	(2.425)
Treatment effect	0.535	-1.806	-1.015
	(3.749)	(1.616)	(3.314)
Lowest self-control			
Control group performance	68.02**	71.69**	58.72**
	(2.947)	(1.062)	(2.566)
Treatment group performance	83.09**	75.99**	67.87**
	(1.985)	(0.751)	(1.552)
Treatment effect	15.07**	4.299**	9.148**
	(2.570)	(0.953)	(2.301)
Observations	446	445	446
Adjusted R^2	0.157	0.107	0.1363
RMSE	18.41	7.138	15.47

Note. Self-control scores are measured by IPIP self-control; robust standard errors in parentheses; full sample results are weighted averages of prefer and not-prefer group.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Table A2. Heterogeneous Treatment Effects by Fail Status

Variable	Class participation rate	iClicker score	Course grade
Did not fail before			
Control group performance	70.9**	72.49**	60.56**
	(2.627)	(0.954)	(2.219)
Treatment group performance	82.05**	75.4**	67.71**
	(1.954)	(0.721)	(1.553)
Treatment effect	11.15**	2.91**	7.145**
	(2.094)	(0.792)	(1.856)
Failed before			
Control group performance	62.16**	71.1**	57.42**
	(4.624)	(1.703)	(4.255)
Treatment group performance	76.62**	73.81**	61.34**
	(3.493)	(1.178)	(2.265)
Treatment effect	14.46**	2.714	3.919
	(5.264)	(1.896)	(4.439)
Observations	453	452	453
Adjusted R^2	0.1495	0.0744	0.1011
RMSE	18.471	7.2236	15.726

Note. Failed before is a dummy variable that indicates whether a student failed a course before; robust standard errors in parentheses; full sample results are weighted averages of prefer and not-prefer group.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$