Higher Education Quality Council of Ontario
An agency of the Government of Ontario

Cooperation and Competition in Large Classrooms

Daniel Brian Krupp[1,2], Joseph Kim[3], Peter Taylor[2] and Pat Barclay[4]

[1]One Earth Future
[2]Queen's University
[3]McMaster University
[4]University of Guelph

Published by

The Higher Education Quality Council of Ontario

1 Yonge Street, Suite 2402
Toronto, ON Canada, M5E 1E5

Phone: (416) 212-3893
Fax: (416) 212-3899
Web: www.heqco.ca
E-mail: info@heqco.ca

Cite this publication in the following format:

Krupp, D. B.[1,2,3], Kim, J.[4], Taylor, P.[3], & Barclay, P.[5] (2014). *Cooperation and Competition in Large Classrooms.* Toronto: Higher Education Quality Council of Ontario.

[1]*Program in Evolution of Governance, One Earth Future*
[2]*Department of Psychology, Queen's University*
[3]*Department of Mathematics and Statistics, Queen's University*
[4]*Department of Psychology, Neuroscience & Behaviour, McMaster University*
[5]*Department of Psychology, University of Guelph*

The opinions expressed in this research document are those of the authors and do not necessarily represent the views or official policies of the Higher Education Quality Council of Ontario or other agencies or organizations that may have provided support, financial or otherwise, for this project.
© Queens Printer for Ontario, 2014

Table of Contents

List of Tables

List of Figures

Acknowledgments

We are exceedingly grateful to Deanna Minervini and Julia Riddell for their help in coordinating this project and for their expert troubleshooting; Greg Atkinson and Chris McAllister for their deft technical support; and the many teaching assistants who made the work possible. We are also indebted to Richard Wiggers, the participants of several HEQCO meetings, and Jill Atkinson, Ingrid Johnsrude, Jami Nelson Nuñez and Eamon Aloyo for comments on the project in its various “moults.”

Executive Summary

Instructors of large classes must contend with numerous challenges, among them low student motivation. Research in evolutionary biology, echoed by work in other disciplines, suggests that aspects of the classroom incentive structure – such as grades, extra credit, and instructor and peer acknowledgment – may shape motivations to engage in studies and to collaborate with peers. Specifically, the way that incentives are distributed in relative quantity (the slope of competition; the proportion of benefits relative to peers associated with performance) and space (the scale of competition; the proportion of peers with whom one is competing) may affect strategies to cooperate or to compete with others.

We hypothesized that students would cooperate with one another more when competition was “global” (i.e., dispersed over the entire population of Introductory Psychology students) than when it was “local” (i.e., concentrated amongst a smaller group of students). We further hypothesized that students would be more motivated when competition was “steep” (i.e., benefits were conditional on relative rather than absolute performance) than when it was “shallow” (i.e., benefits were conditional on absolute rather than relative performance). Moreover, these two variables were expected to interact: cooperation among students was hypothesized to be greatest when competition was both global and steep and weakest when competition was both local and steep.

Here, we designed an experimental test of these hypotheses in a very large, university-level class. Over four semesters, students were randomly assigned, via their tutorial groups, to various competition conditions: global (between-tutorial) competition, local (within-tutorial) competition and asocial (individual) competition. Notably, the global and local competition conditions implied steeper competition than did the asocial competition condition. Within each semester, students were rotated through each condition, so that all students experienced all conditions over distinct testing phases. Students competed over weekly tests for “bonus” credit that could be applied to reweight the course final exam in their favour. We measured their test performance (i.e., scores on the weekly tests) as well as their evaluations of the learning environment (e.g., their reliance on peers and their sense of community in the course).

We predicted that students would perform best and evaluate the learning environment most positively in the global competition condition. However, analyses revealed no such effects. Rather, they simply identified periods of natural variation in test performance (e.g., variation in the difficulty of particular tests) and indicators of a positive learning environment (e.g., variation in students’ reported help given to classmates), irrespective of the experimental manipulation. We identify several possible causes of these null results, including the size of the incentives and possible breakdowns in communication. Despite the null findings, we recommend further experimental work in this area using slightly different incentive schemes and communication systems.

Introduction

The instructor of a large class faces a number of challenges. These might include a motivational crisis among students – who have lesser access to her and greater anonymity than do students in smaller classes – and other such obstacles to student participation and learning (e.g., Bedard & Kuhn, 2008; Bolander, 1973; Finn, Pannozzo & Achilles, 2003; Kokkelenberg, Dillon & Christy, 2008). The problem of motivation is central to the study of behaviour in the social and life sciences of every stripe: economists may ask how employers can optimally motivate their employees; anthropologists may ask why individuals contribute to hunting or communal caregiving; criminologists may ask what makes offenders less likely to recidivate; psychologists may ask why some people have difficulty abstaining from ingesting harmful substances; and biologists may ask why some species exhibit parental care.

In one way or another, scholars have often come to the conclusion that motivation largely turns on incentive structures, though the relationship between incentive and action is anything but obvious (Watts, 2011). From an evolutionary perspective, individuals are expected to be sensitive to tokens of *fitness* (the number of offspring that survive to reproductive age), such as resources, food and mates, because individuals with such sensitivities would have outcompeted those without them (El Mouden, Burton-Chellew, Gardner & West, 2012; Symons, 1987). To the extent that exposure to fitness tokens incentivizes behaviour, we are left with a legacy of incentive-driven motivation: adaptations tailored to help or, alternately, to hinder others when the appropriate incentives are in place.

Fitness-relevant incentives come in a variety of currencies. They may take tangible forms, such as food or money, or they may take other forms entirely, such as reputation and status. In the classroom, incentives can be represented by course marks, depth of knowledge acquired, difficulty of the material, praise from the instructor, and acknowledgment from peers, among many others. In recent years, there has been a push towards new information delivery techniques; this is especially true of the proliferation of “technological” classrooms, where smart boards, interactive clickers and web content are being used to promote learning. Nevertheless, several of the incentives to be found in the classroom are transparently *social*, reflecting a psychology of concern for the effects of performance on both self and others.

Indeed, above and beyond individual concerns, the very motivation to attend and succeed in school is often attributed to short- and long-term goals about *relative* outcomes, as in the ability to pursue a more profitable career, find a good mate, or to earn status and recognition of one’s achievements. Students may desire to outdo their peers in specific content domains (e.g., mathematics, fine arts) or in extracurricular activities (e.g., sports). However, they may also wish to see their allies – friends, romantic partners, teammates, schoolmates – succeed to the extent that their successes are interdependent. Thus, it should be possible to enhance student performance not only by manipulations of material delivery but also by social engagement.

The central argument we make here is that a sufficient change to the social incentive structure of the classroom can inspire better performance in academic study. This requires, of course, some notion of what “sufficient change” might look like. In the current study, this change manifests itself as a manipulation of the distribution of competition.

The Slope and Scale of Competition

Competition is often ill-defined. With whom one competes, to what extent and to what end are elements that are rarely isolated in practice. It is, however, ultimately necessary to do so in order to predict the outcome of a competitive interaction. In terms of social evolution, cooperation and conflict depend on tokens of fitness, as discussed above. But these tokens are distributed both in space and in quantity. This can have profound consequences on behaviour.

We term the distribution of competition in quantity the *slope* or *intensity* of competition. Individuals may compete over resources (e.g., food, mates, status), in which the victors gain a great deal over the vanquished, and so have disproportionately higher fitness. When the stakes are high – that is, when competition is “steep” – it can pay more to invest in conflict than in cooperation. For instance, inequality of income appears to inflame competitive psychology: the homicide rate covaries spectacularly with income inequality at several levels of analysis (municipal, province/state) and eliminates any putative effects of recent cultural history (Daly & Wilson, 2010; Daly, Wilson & Vasdev, 2001; Wilson & Daly, 1997). The upshot is that when the stakes are low – that is, when competition is “shallow” – there is little incentive to engage in conflict with social partners.

More neglected than the slope of competition, however, is the *scale* of it – the distribution of competition in space. Competition may be limited to the individuals directly partaking in a social interaction, but often it is not. Instead, interactions between individuals may have “ripple effects” on others in the population, like dropping a stone into a still pond. Competition is said to be strictly “local” when these ripple effects are constrained to only those individuals directly involved in a social interaction and completely “global” when they affect all others in the population (including those not directly involved) equally.

When competition is local, people are expected to invest in conflict rather than in cooperation (Gardner & West, 2004; Barclay & Stoller, 2014). Yet when competition is global, there can be an incentive to cooperate rather than to compete (West et al., 2006; Barclay & Stoller, 2014). For instance, children are typically local competitors with their siblings over access to parental resources (food, shelter, etc.). However, competition becomes increasingly global as those same siblings mature and begin to look outside of the family for resources. Consequently, they may eventually collaborate with one another. Indeed, there are repeated demonstrations that cooperation within groups increases in relation to competition between them (e.g., Barclay & Benard, 2013; Barclay & Stoller, 2014; Burton-Chellew, Ross-Gillespie & West, 2010; Halevy, Bornstein & Sagiv, 2008; Puurtinen & Mappes, 2009; Radford, 2008; Sherif, Harvey, White, Hood & Sherif, 1954; Van Vugt, De Cremer & Janssen, 2007).

Competition, Cooperation and Classroom Learning

Germane to the work on the effects of competition on cooperation are findings from the field of educational psychology on the effects of cooperation on learning. A meta-analysis of much of this research (Johnson, Maruyama, Johnson, Nelson & Skon, 1981) found that cooperation, including cooperation by intergroup competition, can have substantial benefits for student achievement. For instance, explicit instructions to work together as a group improve student perceptions of teacher support, peer support and test performance relative to instructions to work alone in either competitive or asocial contexts (Johnson, Johnson & Tauer, 1979). Likewise, team-based learning (Michaelsen, Bauman Knight & Fink, 2002) is associated with improvements in student achievement and satisfaction (e.g., McInerney & Fink, 2003; Nieder, Permelee, Stolfi & Hudes, 2005).

Nevertheless, some of this literature is based on methodologically weak designs and much of it is insufficiently grounded in well-supported theories of cooperation. Arguably, the central effect of cooperation will be on motivation – individuals will become motivated to engage with course materials and with their peers more effectively and more deeply. What is now required is careful, rigorous and theoretically sound research into the effects of cooperation on the learning environment.

In spite of work linking cooperation to learning, a great deal of university-level education, especially in large classes, is asocial in its orientation; that is, there is little incentive to cooperate with one's peers. Moreover, some courses that do have social effects are those marked on a de facto "curve," wherein local (i.e., within-class) and steep competition is elicited. The study that we detail below was designed to help evaluate the differential effects of the distribution of competition on several aspects of the learning environment, such as student achievement, student engagement and peer interaction. It is highly rigorous in its design, has reasonable statistical power, can easily be extended to study other cooperative interventions (e.g., "threshold" social dilemmas; Cadsby & Maynes, 1999), and can be readily adapted to other schools, departments and class sizes.

The Current Study

Following the logic above, we argue that altering the distribution of competition in both slope and scale will affect students' motivations to collaborate or compete with their peers. Specifically, we hypothesize that: (1) increasing the slope of competition from shallow to steep will motivate students to dedicate more effort to their studies; and (2) increasing the scale of competition from local to global will motivate students to pitch in with their peers. To the extent that cooperative peer interactions improve learning outcomes, we further hypothesize that (3) students with such assistance will perform better than those without such assistance on course material.

In concrete terms, the manipulation proceeded as follows. First, over four semesters, all students partaking in the Introductory Psychology course at McMaster University competed for bonus credits. These credits were awarded based on student performance and could be redeemed as a re-weighting of final exam scores. For instance, if the original weight of the final exam was 60% of their total course mark and a student earned 3% in credits for his performance, then his final exam would be re-weighted as 57% of his total course mark (should the remainder of his course marks be greater than his final exam mark). A similar credit scheme is used in this course to incentivize students to participate in research studies in the Department of Psychology, Neuroscience and Behaviour at McMaster University, so the students were familiar with bonus credits. Second, criteria for the receipt of bonus credits varied as a function of weekly test performance *within* student tutorials (local competition), *between* tutorials (global competition), or on individual scores *independent* of tutorial performance (asocial competition). The asocial competition condition served as a control for the scale of competition manipulations (global vs. local competition); it also reduced the slope of competition, as all students would have been able to earn bonus credits in this condition, unlike the other conditions (see Methods, below). Notably, all students participated in all competition conditions at some point over the course of the study.

We predicted that students would show the best performance on the weekly tests and evaluate the learning environment most positively when in the global competition condition, relative to their performance in the local and asocial competition conditions. We had no firm predictions regarding performance and evaluation of the learning environment in the local versus asocial competition conditions, because both conditions may provide some positive and some negative motivational components: relative to the asocial competition condition, the local competition condition may undercut within-tutorial cooperation (lowering performance), but its steeper slope of competition may inspire a stronger desire to succeed (raising performance). Finally, we predicted that increased cooperation among students (as measured by the amount of help given and received) would lead to greater test performance, regardless of competition condition.

Methods

The McMaster University Research Ethics Board and the Queen's University General Research Ethics Board approved this study. Participants encompassed all students enrolled in sections of the Introductory Psychology courses at McMaster University in the fall of 2011, winter of 2012, fall of 2012 and winter of 2013. This course follows a "blended learning" model, comprising live lectures, interactive web modules and weekly small-class tutorials (see Sana, Fenesi & Kim, 2011, for additional details). These courses each comprised more than 1,500 students at their smallest enrolment to over 2,500 students at their largest, and each course was subdivided into tutorial groups of (usually) between 10 and 20 students. Students assigned themselves to these tutorial groups based on availability and each tutorial was coordinated by a teaching assistant (TA) who enrolled in an accompanying teacher training course that introduced theory, workshops and practical teaching projects designed to prepare TAs for teaching in the Introductory Psychology course (see Sana, Pachai & Kim, 2011, for additional details). TAs informed their students in tutorial that a study was being conducted on their motivations to learn and how the learning environment affected their success in the course. Students were also informed that the data would be kept anonymously and that their participation in completing the questionnaires was voluntary and would have no effect on their final course mark. This information (including the details of informed consent) was also provided online in each tutorial's web forum, where students could read tutorial-specific information and ask questions of their TAs and fellow tutorial-members.

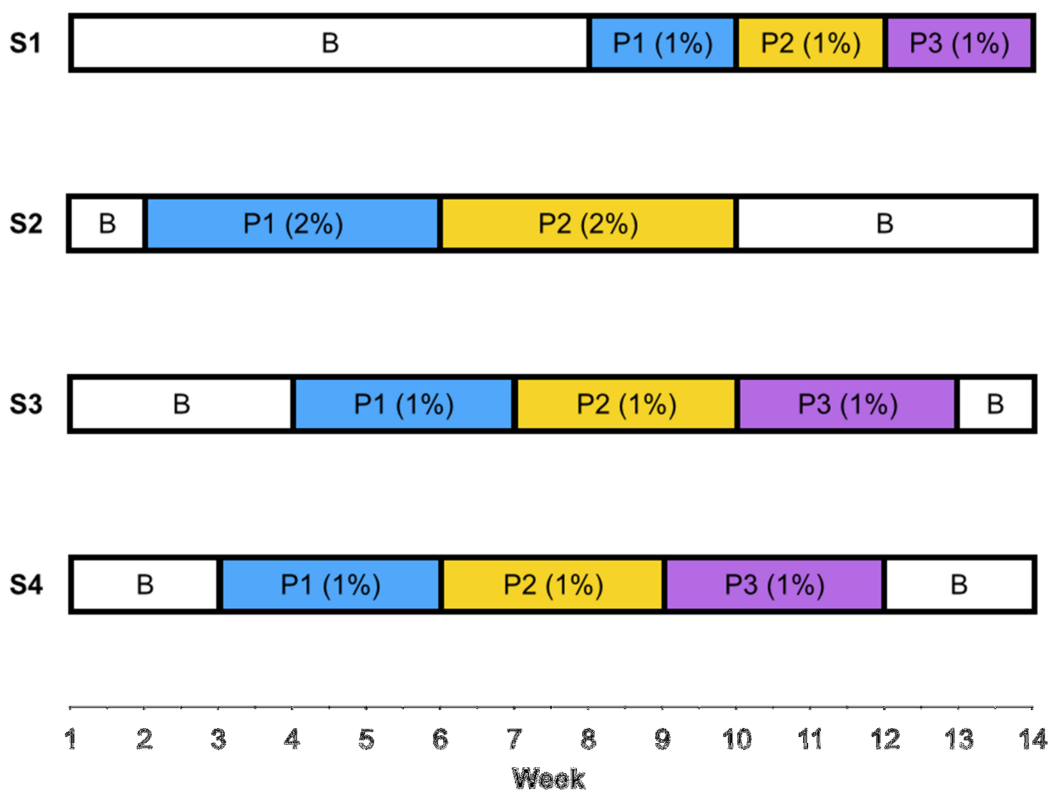
Research Design

Every week over the course of each semester, students were given a short, online test covering the week's material. Students were allowed to take these tests at any point during a 24-hour window and they could do so privately or with the assistance of others; hence, they could work in teams during any of the tests. Our chief dataset comprised the students' scores on these tests. Students in Semesters 3 and 4 were also presented with a modified Course Experience Questionnaire (CEQ; Griffin, Coates, McInnis & James, 2003; McInnis, Griffin, James & Coates, 2001) each week, divided into two parts. The CEQ (Appendix 2) measures aspects of the learning environment, including peer support (e.g., "How many of your classmates helped you to study for or complete this quiz?") and a sense of community (e.g., "My tutorial group feels like a community"). Part I was administered every week, and Part II was additionally administered in the final week of each phase (see below).

Each semester was divided into temporal phases ranging from two to four weeks, which was typically preceded by a baseline period in which no manipulation was introduced. We varied the length and number of these phases as well as the bonus credit rewards across the different semesters, so the manipulation and results for each semester are presented separately below. Figure 1 represents the timelines and the key variations in the study over the four semesters. As mentioned previously, all students were assigned to a tutorial group coordinated by a TA, most of whom were responsible for two or three tutorial groups. Because we could not have some students assigned to one competitive condition but others within the same tutorial group assigned to another condition, we assigned all students of a given TA to a particular order of condition presentation. Thus, TAs were randomly assigned to a condition for the first phase and then assigned again to one of the remaining conditions in each of the remaining phases. Consequently, students were exposed to each of the local, global and asocial competition conditions, assigned to a particular order via their TA. In all semesters, TAs informed their students of the structure of the competitive manipulations using a script (see Appendix 1 for an example) that was emailed to them at the onset of each phase. This script was read aloud to the students during the tutorial and then posted online in the tutorial's web forum. Students were neither given the hypotheses nor the labels we use here to explain them (e.g., "global competition"). In each phase, TAs were asked to confirm by email that they had read the script to their students and posted it in their tutorial's online forum.

In Semester 1 (S1; fall 2011), students were exposed to global, local and asocial competition conditions. Each phase consisted of one two-week testing block (phase 1, tests 8 and 9; phase 2, tests 10 and 11; phase 3, tests 12 and 13), where students were told that they could reduce the weight of their final exam by up to 3% (Figure 1). TAs were randomly assigned to a condition in the first phase and then nonrandomly assigned in the subsequent phases to the remaining two conditions: TAs who began in the global competition condition were assigned to asocial competition in phase 2 and local competition in phase 3; those assigned to local competition in phase 1 were assigned to global competition in phase 2 and asocial competition in phase 3; and those assigned to asocial competition in phase 1 were assigned to local competition in phase 2 and global competition in phase 3.

Figure 1: Timeline of the Manipulation



In each semester (**S1** to **S4**), students were assigned via their TAs to the experimental conditions, each lasting from two to four weeks. Following a baseline period (**B**) at the onset of each of each semester, students were assigned to one condition during the first phase (**P1**) then reassigned to the remaining conditions during the second (**P2**) and third (**P3**) phases. (Note: there were only two conditions and, hence, two phases in Semester 2.) The value of the bonus credits available in each phase is given in parentheses.

In the global competition condition, students were informed that if their tutorial group's average mark over the two quizzes in the current testing phase was in the top 25% of average marks across all tutorials over the same period, they would receive 1% credit. In the local competition condition, students were told that if their own average mark on the two quizzes in the current testing phase was in the top 25% of average marks in their tutorial over the same period, they would receive 1% credit. Finally, in the asocial competition condition, students were told that if their average mark on the two quizzes in the current testing phase was 9 out of 10 or above, they would receive 1% credit. Thus, students could earn up to 3% credit for reweighting the final exam over the course of the three study phases in S1. As the phases progressed, students who had earned credit were emailed to provide them with this information and to congratulate them.

In Semester 2 (S2; winter, 2012), we decided to limit the manipulation to just the global and local competition conditions, so that students would be exposed to the experimental conditions for longer periods of time. Consequently, S2 TAs (and their students) were randomly assigned to a condition in the first phase of the study and were subsequently assigned to the remaining condition in the second phase. Each phase consisted of one four-week testing block (phase 1, tests 2 to 5; phase 2, tests 6 to 9), where students were told that they could reduce the weight of their final exam by up to 4% (Fig. 1). Students were given similar instructions as in S1, but were informed that they could earn 0.5% credit if their tutorial's average mark was in the top 25% across all tutorials over each test in the global competition condition phase (for a total stake of 2% credit) and another 0.5% credit if their own mark was in the top 25% of their tutorial over each test in the local competition condition phase (for a total stake of another 2% credit). In S2, students who earned credit were emailed each week to provide them with this information and to congratulate them.

In Semester 3 (S3; fall 2012), we returned to three competition conditions (global, local and asocial) but allowed the phases to run for three weeks rather than two (as in S1) or four (as in S2). In S3, TAs were randomly assigned to a condition, without replacement, in each phase. Students could again earn up to 3% credit for reweighting their final exams, partitioned into 1% units for each testing phase (phase 1, tests 4 to 6; phase 2, tests 7 to 9; phase 3, tests 10 to 12; Figure 1). Instructions for the competition conditions were the same in structure as in S1. Students were only informed of their credit earnings after the testing phases concluded. In S3, we also introduced the CEQ. Prior to completing the online tests, students were required to complete a pre-test. As part of this pre-test, students were given the short form of the CEQ (Part I only) for the first two weeks of each phase and the long form (Parts I and II combined) for the final week of each phase.

Finally, Semester 4 (S4; winter 2013) followed the same design as S3: all three competition conditions were tested, each for one three-week phase (phase 1, tests 3 to 5; phase 2, tests 6 to 8; phase 3, tests 9 to 11) and each worth 1% credit for a total of 3% reweighting of the final exam (Figure 1). Again, the TAs were randomly assigned to a condition (without replacement) in each of the phases. The instructions matched those of S1 and S3, students were only informed of their credit earnings after the testing phases concluded, and students were asked to complete the short form of the CEQ for the first two weeks of each phase and the long form for the last week of each phase.

Results

Analytical Approach

Recall that the TAs were randomly assigned to a condition for the first phase in S1 and for each testing phase in S2-S4, and so their tutorial students were, as a group, assigned to these same conditions over the phases. However, the data collected pertain to student rather than TA performance. Thus, an analysis of the student data risks pseudoreplication (Hurlbert, 2009) – the experimental unit is at the level of the TAs but the evaluation unit is at the level of the students. To manage this, we approached the data with a nested, mixed-factorial analysis of variance (ANOVA).

The independent variables (i.e., the variables that we manipulated) for the study were “Condition” (global, local and asocial competition), “Order” (the order in which each student was presented with each condition over the testing phases) and “TA” (the identity of the TA responsible for each student’s tutorial group). There were three different levels of Condition in S1, S3 and S4 and two levels of this variable (global and local competition) in S2. As TAs were randomly assigned to a condition in only the first phase in S1 but in all phases in S2-S4, there were three different, between-subject orders of the conditions by testing phase in S1, two in S2, and six in S3 and S4. Finally, there were as many levels of TA as there were TAs in the study over a given semester. Although TAs may affect their students’ performance, we were only interested in the TA variable so that we could evaluate student test scores and CEQ responses at the appropriate level of analysis. To do so, we nested the TA variable within the Order variable, because TAs represent random subgroups within the Order manipulation. Condition was a within-subjects variable, because all TAs and their students were exposed to all competition conditions. Order and TA were between-subjects variables, because TAs and their students were only exposed to one order of the conditions over the testing phases and students were assigned to only one TA over the course of the study.

The dependent variables for S1 through S4 were individual students’ scores on the weekly tests. In addition, individual students’ responses (averaged separately for each phase) from S3 and S4 to the three items from Part I of the CEQ (“How many of your classmates did you help to study for or complete this quiz?”, “How many of your classmates helped you to study for or complete this quiz?” and “Did you use your course materials while completing the quiz?”) and the total score on Part II of the CEQ were analyzed in separate models. For ease of comprehension, we summarize the statistical effects here, but the full results are available in Appendix 3. Note that we report our findings to be statistically “significant” using the conventional requirement of our tests being associated with a p value of less than .05.

It is possible that the competition manipulation had the strongest (or sole) effects in the initial phase of the study, when it was first introduced. We tested this by analyzing the effects of the independent variables on test scores in the first phase of each semester alone. Thus, we analyzed effects of Condition and TA (nested within Condition) on student performance; both were between-subjects variables in these models.

Finally, we tested whether there was any effect of help given, help received, or CEQ Part II responses on test scores. Restricting the analyses to the weeks of the study manipulation, we ran multiple regressions predicting mean test score from the mean amounts of help given (S3 and S4), mean amounts of help received (S3 and S4), and CEQ Part II scores (S3 only).

Findings by Semester

In S1, there were no significant main effects of any of the independent variables on test scores. However, there was a significant interaction between Condition and Order. As can be seen from Table 1, which relates these two variables, students performed significantly better in the condition to which they were exposed in phase 3 than in the other two phases. They also performed significantly better in the condition to which they were exposed in phase 2 than in the condition to which they were exposed in phase 1. This suggests that the interaction between Condition and Order is the result of changes in test performance over the phases of the study rather than an effect of the manipulation itself. In sum, we did not find the predicted effect of increased test performance in the global competition condition relative to the local and asocial competition conditions.

Table 1: Rank of Test Scores by Condition in Relation to their Order of Appearance, S1

Order of appearance of conditions	Rank of test scores by condition
1. Global	1. Local
2. Asocial	2. Asocial
3. Local	3. Global
1. Local	1. Asocial
2. Global	2. Global
3. Asocial	3. Local
1. Asocial	1. Global
2. Local	2. Local
3. Global	3. Asocial

In S2, there were also no significant main effects of any of the independent variables on test scores. However, there was again a significant interaction between Condition and Order. As can be seen from Table 2, students performed significantly better in the condition to which they were exposed in phase 1 than in the condition to which they were exposed in phase 2. This again suggests that the interaction between Condition and Order is the result of changes in test performance over the phases of the study rather than an effect of the manipulation. In sum, we did not find the predicted effect of increased test performance in the global competition condition relative to the local competition condition.

Table 2: Rank of Test Scores by Condition in Relation to their Order of Appearance, S2

Order of appearance of conditions	Rank of test scores by condition
1. Global	1. Global
2. Local	2. Local
1. Local	1. Local
2. Global	2. Global

In S3, there were also no significant main effects of any of the independent variables on test scores. Yet again, however, there was a significant interaction between Condition and Order. As can be seen from Table 3, students performed significantly better in the condition to which they were exposed in phase 3 than in the conditions to which they were exposed in phases 1 and 2. Once more, this suggests that the interaction between Condition and Order is the result of changes in test performance over the phases of the study rather than an effect of the manipulation.

Table 3: Rank of Test Scores by Condition in Relation to their Order of Appearance, S3

Order of appearance of conditions	Rank of test scores by condition
1. Global 2. Asocial 3. Local	1. Local 2. Global & Asocial
1. Global 2. Local 3. Asocial	1. Asocial 2. Global & Local
1. Local 2. Global 3. Asocial	1. Asocial 2. Global & Local
1. Local 2. Asocial 3. Global	1. Global 2. Local & Asocial
1. Asocial 2. Local 3. Global	1. Global 2. Local & Asocial
1. Asocial 2. Global 3. Local	1. Local 2. Global & Asocial

There were no significant main effects of or interactions among any of the independent variables on the amount of help given. However, there were significant interactions between Condition and Order on amount of help received, frequency of material use, and CEQ Part II scores. Inspection of the simple main effects suggests that students tended to: (1) receive more help in the global competition condition than in the other two conditions, but only when global competition was encountered first; (2) use course materials more frequently during certain phases of the study than during others; and (3) consider the learning environment more positively during certain phases of the study than during others. The latter two findings are in keeping with those regarding the test scores, namely, that the Condition by Order interactions are largely indicative of changes in frequency of course material use and assessment of the learning environment over time rather than of a consequence of the competition manipulation *per se*. In sum, we did not find the predicted effect of increased test performance or improvements to the learning community (with the possible exception of the amount of help received) in the global competition condition relative to the local and asocial competition conditions.

In S4, there were two unfortunate errors that necessitated more restrictive data analysis. First, the online platform designed to deliver the tests failed for test 3 and the associated CEQ items. Consequently, we limited analysis of the test and CEQ data in the first phase of the semester to the weeks of tests 4 and 5, rather than to the weeks of tests 3, 4 and 5. Second, the delivery of the CEQ was not correctly aligned with the phases: Part I was presented in all weeks, as planned, but Part II of the CEQ was administered during the weeks of tests 6, 9 and 12, when it should have been administered during the weeks of tests 5, 8 and 11. Thus, Part II of the CEQ was administered at the beginning of phases 2 and 3 and after the manipulations had concluded rather than being administered at the end of phases 1, 2 and 3. Because analysis of Part II of the CEQ would confuse different phases, we limited analysis to Part I of the CEQ only.

There was a significant main effect in S4 of TA nested within Order on test scores, but no other main effects of any of the other independent variables. Yet again, however, there was a significant interaction between Condition and Order. As can be seen from Table 4, students performed significantly better in the condition to which they were exposed in phases 2 and 3 than in the conditions to which they were exposed in phase 1. For students in Order 5, performance in phase 3 (global competition) was also significantly greater than in phase 2 (local competition). This aside, the pattern of results again suggests that the interaction between Condition and Order is the result of changes in test performance over the phases of the study rather than an effect of the manipulation.

Table 4: Rank of Test Scores by Condition in Relation to their Order of Appearance, S4

Order of appearance of conditions	Rank of test scores by condition
1. Global 2. Asocial 3. Local	1. Asocial & Local 2. Global
1. Global 2. Local 3. Asocial	1. Asocial & Local 2. Global
1. Local 2. Global 3. Asocial	1. Global & Asocial 2. Local
1. Local 2. Asocial 3. Global	1. Global & Asocial 2. Local
1. Asocial 2. Local 3. Global	1. Global 2. Local 3. Asocial
1. Asocial 2. Global 3. Local	1. Global & Local 2. Asocial

There were no significant main effects of or interactions among any of the independent variables on the amount of help given or the amount of help received. However, there was a significant interaction between Condition and Order on the frequency of material use. Inspection of the simple main effects suggests that students tended to use course materials more frequently during certain phases of the study than during others. The findings are again in keeping with those regarding the test scores: the Condition by Order interactions are indicative of changes in frequency of course material use over time rather than of a consequence of the competition manipulation. In sum, we did not find the predicted effect of increased test performance, cooperation or improvements to the learning community in the global competition condition relative to the local and asocial competition conditions.

Analysis of First Phase Test Scores

There were no significant effects of or interactions between either of the independent variables on test score in the first phases of S1, S2 or S3. There was, however, a significant effect of TA in the first phase of S4 (restricted to the data from tests 4 and 5; see above). In sum, we did not find the predicted effect of increased test performance in the global competition condition relative to the local and asocial competition conditions.

Effects of Help Given and Received

There were no significant effects of mean amounts of help given or help received on mean test scores in S3 or S4 (restricted to the data from tests 4 and 5; see above). However, mean CEQ Part II scores significantly and positively predicted mean test scores in S3. In sum, we did not find a positive association between test performance and help given or received, though we did find a positive association between a measure of a positive learning community and test performance.

Discussion

Summary of Findings

The study was designed to assess the effects of the scale and slope of competition on student performance and the learning environment in large classes. We used an experimental design in which thousands of university students were randomly assigned (via their TAs) to one of several competition conditions – instantiated by competition between tutorials (global competition), within tutorials (local competition), or with oneself (asocial competition) – and subsequently rotated through the remaining conditions. Moreover, we replicated the essential elements of the design four times. The primary hypothesis was that global (and steep) competition would reinforce cooperation among students within tutorial groups and that this in turn would foster better performance on the weekly tests and a more positive learning environment.

Overall, the findings of our study are inconsistent with this hypothesis. We found no effect of global competition on test performance (across all phases or within the first phase alone) or on responses to any of the items on the CEQ in any of the four semesters. Moreover, we did not find any effect of amounts of help given or received on test performance, though we did find a positive association between a measure of the learning environment (mean CEQ Part II responses) and student performance on the weekly tests. This suggests that cooperation can improve student performance, though it is difficult to draw firm conclusions in this regard, as the direction of the causal relationship between the learning environment and test performance cannot be determined from our design.

We also found, on numerous occasions, interactions between competition condition and the order in which the conditions were presented to the students. However, closer examination of these effects indicates that patterns of variation in performance and CEQ responses were caused by variation in test difficulty (or student preparedness) over the semester, rather than by changes in experimental condition. For instance, students generally performed best on tests in the last phase of the study in S1, in the first phase of the study in S2, in the last phase of the study in S3, and in the last two phases of the study in S4. Similar effects were found for the Condition by Order interactions for CEQ items in S3 and S4, such as the finding that students used their course materials more frequently in the earliest phase of the study in both semesters.

There is one notable exception to the above pattern of results. In S3, students tended to receive more help in the global competition condition than in the local or asocial competition conditions. However, this effect only materialized when global competition was encountered first. Moreover, help received during global competition was only significantly greater than help received during asocial competition (but not during local competition) in one order and only significantly greater than help received during local competition (but not during asocial competition) in another order. Finally, this effect of global competition did not replicate in S4, which was very similar in design to S3. All of this suggests that the effect of global competition on help received is small, at best, and quite possibly illusory.

Limitations and Future Directions

The failure to support the hypothesis that global competition reinforces cooperation within tutorials and subsequent performance on tests may stem from a number of sources. First, and most obviously, the hypothesis may be incorrect. While certainly possible, this seems unlikely to us. There is a substantial literature on the positive effects of global or between-group competition on within-group cooperation (e.g., Barclay & Benard, 2013; Barclay & Stoller, 2014; Burton-Chellew, Ross-Gillespie & West, 2010; Halevy, Bornstein & Sagiv, 2008; Puurtinen & Mappes, 2009; Radford, 2008; Sherif, Harvey, White, Hood & Sherif, 1954; Van Vugt, De Cremer & Janssen, 2007), suggesting that the finding is relatively robust. It is possible that, in the classroom, the effect may be quite small. Nevertheless, even if the effect of global competition on cooperation were small, it is difficult to deny that we had a sufficiently large sample size to detect it.

Second, the failure to support the hypothesis may reflect the size of the incentives, credits that would have had at most a 4% effect on the weighting of the final exam. However, we are again sceptical of this explanation. Students seem very motivated by these credits in the context of research participation, where the opportunity costs per credit are likely higher. For example, students must spend at least one hour in the laboratory per credit, whereas they need not spend a moment longer collaborating on weekly tests with their peers and would also see their marks improve. With that said, it would be simple enough to increase the size of the incentive or to find a different kind of incentive (including more intrinsic rewards) in future work. For example, students may respond better to the threat of losing a bonus (i.e., unearned) credit for failing to perform according to the standards of the manipulation than to the opportunity to gain such a credit (Kahneman, Knetsch & Thaler, 1991; Kahneman & Tversky, 1984). Likewise, students may value reputational benefits (e.g., public recognition within the tutorial group or class) more than minor adjustments to grades. Relatedly, it is also possible that the criteria of scoring in the top 25% of all students within each tutorial or in the top 25% of all tutorial groups is not sufficient to inspire competition among the students. Hence, it might be possible to produce effects when the proportion of possible winners is reduced.

Third, there may have been a breakdown in communication between the researchers and the students with respect to implementation. On several occasions, students contacted the researchers, confused about the details of the experimental conditions, about the awarding of credits, or about their performance. Indeed, some students expressed surprise that they were participating in a study at all, despite being repeatedly provided with information about it.

The instructions were to be relayed by the TAs to their students, but it was rare to get timely confirmation of this. In fact, many TAs had to be reminded more than once to confirm that they had provided the instructions to their students in tutorial and online. Moreover, students may not have attended tutorial, paid attention to the instructions, or remembered the instructions when they were relevant. Thus, it is possible that the basic experimental design was adequate, but the method of information delivery was insufficient. This strikes us a likely explanation for our results.

Finally, much of the research on cooperation in the classroom explicitly emphasizes the communal aspects of peer interaction, actively encouraging group work, framing it in the form of “teamwork,” and providing rewards for working together (e.g., Johnson, Johnson & Tauer, 1979; Michaelsen, Bauman Knight & Fink, 2002). Part of our interest in a “minimalist” design – manipulating the scale and slope of competition with credit incentives alone – was to offer a method that required little effort on the part of instructors teaching large classes, because it might be otherwise difficult to actively encourage cooperation in such classes. Nonetheless, there is undoubtedly some middle ground between an incentive structure that has been stripped bare of its social context and one that relies on a heavy dose of contextual information. Future research should study this middle ground more thoroughly.

Conclusions

The results of our study do not support the hypothesis that global competition fosters within-tutorial cooperation or student performance. This does not mean that the hypothesis is incorrect (see above), but it does suggest that more will have to be done to produce such effects, whether that means providing stronger incentives, solving the problem of communication between researchers and students, or steeping the incentive structure in social context.

Although the findings of the study were, by and large, null, we hope that the work will inspire further research into the problem of competition, cooperation and student motivation and into experimental methods in the classroom. In developing the work, we were often approached by colleagues who cited difficulties with or a lack of appreciation for experimental methods in applied settings. Yet we were able to implement a study of thousands of students, randomly assigned to condition, over the course of four semesters. This design has many virtues. First, it can address cause and effect in ways that less rigorous designs cannot. Second, this design is much less susceptible to “demand” characteristics, whereby students behave in particular ways because they have been led to believe (as an accidental consequence of the study) that they *ought* to behave in such ways, and experimenter bias, whereby the researchers can unintentionally influence study outcomes. Evidence of robustness to demand characteristics comes from the temporal patterns of test scores and responses to the CEQ: students did not reliably perform best or report a better learning environment at the onset of the study from one semester to the next.

To be fair, there are numerous obstacles to implementing experimental methods in the classroom and not all of them are surmountable. So let us end with the following advice: an experimental approach is preferable when possible; when it is not, there are many excellent quasi-experimental methods that could do the trick (e.g., Shadish, Cook & Campbell, 2001). To the extent that this advice is followed, we expect to see many successful solutions to the problems posed by large classes on the horizon.

References

Barclay, P., & Benard, S. (2013). Who cries wolf, and when? Manipulation of perceived threats to preserve rank in cooperative groups. *PLOS ONE*, *8,* e73863.

Barclay, P., & Stoller, B. (2014). Local competition sparks concerns for fairness in the ultimatum game. *Biology Letters, 10,* article 20140213.

Bedard, K., & Kuhn, P. (2008). Where class size really matters: Class size and student ratings of teacher effectiveness. *Economics of Education Review, 27,* 253-265.

Bolander, S. F. (1973). Class size and levels of motivation. *Journal of Experimental Education, 42,* 12-17.

Burton-Chellew, M. N., Ross-Gillespie, A., & West, S. A. (2010). Cooperation in humans: competition between groups and proximate emotions. *Evolution and Human Behavior, 31*, 104-108.

Cadsby, C. B., & Maynes, E. (1999). Voluntary provision of threshold public goods with continuous contributions: experimental evidence. *Journal of Public Economics, 71,* 53-73.

Daly, M., & Wilson, M. (2010). Cultural inertia, economic incentives, and the persistence of "Southern violence". In M. Schaller, A. Norenzayan, S. Heine, T. Yamagishi & T. Kameda (eds.), *Evolution, Culture, and the Human Mind* (pp. 229-241). New York: Taylor & Francis.

Daly, M., Wilson, M., & Vasdev, S. (2001). Income inequality and homicide rates in Canada and the United States. *Canadian Journal of Criminology, 43*, 219-236.

El Mouden, C., Burton-Chellew, M., Gardner, A., & West, S. A. (2012). What do humans maximise? In S. Okasha & K. Binmore (eds.), *Evolution and Rationality* (pp. 23-49). Cambridge: Cambridge University Press.

Finn, J. D., Pannozzo, G. M., & Achilles, C. M. (2003). The “why’s” of class size: Student behaviour in small classes. *Review of Educational Research, 73,* 321-368.

Gardner, A., & West, S. A. (2004). Spite and the scale of competition. *Journal of Evolutionary Biology, 17*, 1195-1203.

Griffin, P., Coates, H., McInnis, C., & James, R. (2003). The development of an extended course experience questionnaire. *Quality in Higher Education, 9,* 259-266.

Halevy, N., Bornstein, G., & Sagiv, L. (2008). “In-group love” and “out-group hate” as motives for individual participation in intergroup conflict: a new game paradigm. *Psychological Science, 19,* 405-411.

Hurlbert, S. H. (2009). The ancient black art and transdisciplinary extent of pseudoreplication. *Journal of Comparative Psychology, 123,* 434-443.

Johnson, R. T., Johnson, D. W., & Tauer, M. (1979). The effects of cooperative, competitive, and individualistic goal structures on students’ attitudes and achievement. *Journal of Psychology, 102,* 191-198.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin, 89*, 47-62.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives, 5,* 193-206.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39,* 341-350.

Kokkelenberg, E. C., Dillon, M., & Christy, S. M. (2008). The effects of class size on student grades at a public university. *Economics of Education Review, 27,* 221-233.

McInerney, M., & Fink, L. D. (2003). Team-based learning enhances long-term retention and critical thinking in an undergraduate microbial physiology course. *Journal of Microbiology & Biology Education, 4,* 1-18.

McInnis, C., Griffin, P., James, R., & Coates, H. (2001). *Development of the Course Experience Questionnaire (CEQ)*. Canberra: Department of Education, Training and Youth Affairs.

Michaelsen, L. K., Bauman Knight, A., & Fink, L. D. (eds.) (2002). *Team-based learning: A transformative use of small groups*. Westport, CT: Praeger.

Nieder, G. L., Permelee, D. X., Stolfi, A., & Hudes, P. D. (2005). Team-based learning in a medical gross anatomy and embryology course. *Clinical Anatomy, 18,* 56-63.

Puurtinen, M., & Mappes, T. (2009). Between-group competition and human cooperation. *Proceedings of the Royal Society, Series B: Biological Sciences, 276,* 355-360.

Radford, A. N. (2008). Duration and outcome of intergroup conflict influences intragroup affiliative behaviour. *Proceedings of the Royal Society, Series B: Biological Sciences, 275,* 2787-2791.

Sana, F., Fenesi, B., Kim, J. A. (2011). A Case Study of the Introductory Psychology Blended Learning Model. *Canadian Journal of Scholarship of Teaching and Learning, 2,* 1-10.

Sana, F., Pachai, M., Kim, J. A. (2011). Role of training for undergraduate teaching assistants in a peer mentor course. *Transformative Dialogues, 4,* 1-10.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Second ed. Stamford, CT: Cengage.

Sherif, M., Harvey, O. J., White, J., Hood, W. R., & Sherif, C. W. (1954). *Intergroup conflict and cooperation: The Robbers Cave experiment*. Norman, OK: University of Oklahoma.

Symons, D. (1987). If we're all Darwinians, what's the fuss about? In C. Crawford, M. Smith & D. Krebs (eds.), *Sociobiology and psychology* (pp. 121-146). Hillsdale, NJ: Erlbaum.

Van Vugt, M., De Cremer, D., & Janssen, D. P. (2007). Gender differences in cooperation and competition: the male-warrior hypothesis. *Psychological Science, 18,* 19-23.

Watts, D. J. (2011). *Everything is obvious*. New York: Crown.

West, S. A., Gardner, A., Shuker, D. M., Reynolds, T., Burton-Chellow, M., Sykes, E. M., et al. (2006). Cooperation and the scale of competition in humans. *Current Biology, 16*, 1103-1106.

Wilson, M., & Daly, M. (1997). Life expectancy, economic inequality, homicide, and reproductive timing in Chicago neighbourhoods. *British Medical Journal, 314*, 1271-1274.

Higher Education Quality Council of Ontario

An agency of the Government of Ontario