Does the Response to Competition Depend on Perceived Ability? Evidence from a Classroom Experiment *

Kelly Bedard
University of California, Santa Barbara and IZA
kelly.bedard@ucsb.edu

Stefanie Fischer
Cal Poly State University, San Luis Obispo
sjfische@calpoly.edu

February 5, 2019

Abstract

We examine the effect of relative evaluation on test performance by implementing a classroom-level experiment in which students are financially incentivized either individually or in a tournament. Linking the experimental data with student-level administrative data allows us to study two aspects of competitive environments: tournament structure and one's perceived position in the ability distribution. At least in our setting, we find only limited evidence that effort responses to competition are sensitive to tournament size or prize structure. However, in contrast to previous studies that examine effort responses to exogenously assigned competition, we find a large negative competition effect for students who believe they are relatively low in the ability distribution, and no competition effect for those who believe they are relatively high ability. Using additional treatments, we further show that the divergence between our results and past results is likely driven by task type and not by differences in selection into participation between lab and classroom environments.

*We thank the office of Institutional Research, Planning and Assessment at UC Santa Barbara for access to administrative data. We also thank the UCSB economics and accounting instructors who donated class time to the experiment. This paper has benefited from helpful comments from Jenna Stearns, Emanuel Vespa, Ryan Oprea, Gary Charness, Aric Shafran, Brian Duncan, Rey Hernandez-Julian, and seminar participants at the Federal Reserve Bank of New York and Claremont McKenna College. This experiment and research has been conducted with IRB approval from the University of California, Santa Barbara. All errors are our own.

1 Introduction

Education, employment, athletics and many other settings feature relative payment schemes in the forms of tournaments, bonus payments and grading curves. There are empirical and experimental literatures documenting differential effort responses to such competition by gender, tournament size and payment structure.[1] Less is known about differences in responses across the ability/skill distribution.[2] We contribute to the existing literature by exploring differential responses to tournament size and payment structure across students with different perceptions about their ability/skill using an effort-based classroom experiment. Our setting is well-suited for this purpose because undergraduates have at least some knowledge about their mastery of a subject relative to their cohort as they receive grades on a regular basis. To be clear, effort refers to immediate effort exerted on a financially incentivized economics quiz that does not count towards a student's course grade.

We are, of course, not the first to study effort responses to forced competition. In most cases, effort is either modeled as a cost chosen off a menu in a laboratory setting (Bull et al., 1987; Orrison et al., 2004) or as actual effort on a simple task such as solving mazes, running, adding up numbers, or solving word and memory puzzles (Gneezy et al., 2003; Gneezy and Rustichini, 2004; Günther et al., 2010; Dreber et al., 2014). These studies generally find that individuals who are forced to compete work at least as hard as those who are not, though there is some evidence suggesting the competition effect is sensitive to task type.[3]

Ability likely plays an important role in effort response. Brown (2011) offers an intuitive model showing how ability gaps (or perceived gaps) between competitors can result in reduced effort for

[1] Gender differences are shown by Gneezy et al. (2003) and Günther et al. (2010) in lab experiments, Gneezy and Rustichini (2004) and Gneezy et al. (2009) in field experiments, and Fischer (2017) in STEM major persistence. Tournament size is explored by Barut et al. (2002), Orrison et al. (2004), List et al. (2014) and Lim et al. (2014), and payment structure by Harbring and Irlenbusch (2003), Orrison et al. (2004) and Chen et al. (2011).

[2] Evidence of differential responses by ability are documented by Müller and Schotter (2010), Brown (2011), Jalava et al. (2015), Andreoni and Brownback (2017) and Brownback (2018).

[3] There is a closely related experimental literature examining the propensity of individuals to choose competition. This literature finds that women are less likely to choose competition over a piece rate option compared to men (Niederle and Vesterlund, 2007; Cason et al., 2010; Sutter and Glätzle-Rützler, 2014; Booth and Nolen, 2012; Villeval, 2012; Datta Gupta et al., 2013; Garratt et al., 2013). There also exists an extensive theoretical literature focused on understanding effort response to relative performance incentive schemes, examples include Lazear and Rosen (1981), Green and Stokey (1983), and Prendergast (1999). For a comprehensive review of the experimental literature on competition – including theory, lab and field – see Dechenaux et al. (2015).

weaker contestants. She provides empirical support for her model using data from professional golf. More specifically, she shows that competitors perform worse when Tiger Woods plays in a tournament, at least in his prime. As such, it seems likely that people with different ability levels respond differentially to competition in other settings as well. There is a related empirical literature examining the impact of grading standards on effort, studying, test scores, grades and educational attaintment. Oettinger (2002), Betts and Grogger (2003) and Figlio and Lucas (2004) all find that higher grading standards increase effort and test scores, but that the effects are unequally distributed across students.[4]

Economic theory offers important insights about the range of possible effort responses to tournament size. For example, List et al. (2014) show that whether effort increases, decreases or remains the same as tournament size grows, depends on the distribution of luck associated with particular environments. Andreoni and Brownback (2017) make an important contribution to this literature by incorporating heterogeneity in ability.[5] In this framework, students become less certain about their rank as class size shrinks. This uncertainty has asymmetric effort effects; increasing effort at the bottom of the distribution and decreasing it at the top. One key take away from this literature is that effort response to tournament size is likely context specific because the distributions of idiosyncratic luck and ability vary.

Payment structure coupled with an ability distribution likely has implications for effort response. The Chen et al. (2011) model shows that as the number of winners increase, high types exert at most as much effort while low types exert at least as much effort. Intuitively, effort is ultimately determined by the marginal probability of winning. Predictions from this model highlight the tradeoff that must be made when increasing number of winners; a possible reduction in effort from high types in favor of a possible increase in effort from lower types.

While models help build intuition, in environments outside the lab where there is less control over the parameters of the experiment, predicting effort responses is difficult. In other words,

[4]There is a larger literature examining the ability of financial incentives to increase effort. Recent examples include, Angrist et al. (2009), Angrist and Lavy (2009), Kremer et al. (2009), Leuven et al. (2010), Fryer Jr (2011) and Burgess et al. (2016).

[5]Their model builds on work by Lazear and Rosen (1981), Becker and Rosen (1992) and Orrison et al. (2004).

student effort responses to competition, and whether they vary systematically by ability and tournament structure are in the end empirical questions. Our objective is to examine how effort changes in response to competition in a classroom setting, and to further ask whether responses differ across ability groups and/or the size and payment structure of the tournament in which one is forced to compete. To do this, we run experiments at the beginning of economics classes early in the academic quarter. Students are randomly assigned across a variety of tournament structures in which their earnings depend on their own quiz performance relative to their opponent's performance. To isolate the causal effect of competition, we also include a non-competitive piece rate treatment in which subjects are paid a flat rate for each correct answer.

We find a large negative competition effect for students who believe they are a relatively low scoring student. In particular, students who believe they will earn less than an "A" grade in their current economics course score 16-33% of a standard deviation lower when forced to compete compared to students assigned to the piece rate treatment.[6] We will refer to these students as low expectation students. In contrast, there is no statistically significant difference in quiz scores between tournament and piece rate subjects for those who believe they will earn an "A". We will refer to these students as high expectation students. Consistent with results from a lab experiment presented in Lim et al. (2014), we find at best limited evidence that these results vary across tournament size.[7] We also find no evidence that they vary by payment structure.

There are two obvious candidate explanations for the divergence between our results and those reported in previous studies: task type and selection into participation. We explore these possibilities by running two additional treatments. In the first we replace the microeconomics quiz with a simple numeric task that is more similar to previous experiments. Interestingly, in these rounds we find *no* difference across competitive and piece rate payment schemes for either low or high expectation students – the result that weaker students exert less effort in competition vanishes. In

[6]For context, 62% of students believe that they will earn an A or an A- (what we refer to as an "A" grade), despite the fact that the actual number of such grades awarded is far below this level. We explore the difference between expected grades and actual earned grades in Section 5.1.

[7]More specifically, only large tournaments under one payment structure elicits less effort for low expectation students. In contrast, in a lab experiment Andreoni and Brownback (2017) show high valuation types exert more effort in larger groups while low types reduce effort.

the second additional treatment we move the experiment (using the same microeconomics quiz) to the end of class/section and tell students that they are free to stay and participate or to leave. Approximately fifty percent of students choose to leave. The results for these rounds are similar to the rounds that are run at the beginning of class/section. In other words, we find no evidence that our competition results are driven by selection into participation.

The findings in this study offer at least two new insights. First, in our classroom context where students are asked to complete a real-effort and ability-specific task, we show responses to competition depend on perceived ability. When forced to compete, low expectation students work less hard and high expectation students do not change their effort level. Second, the fact that results from rounds of the experiment in which a more general skill task is used – one similar to those used in previous studies – differ from the results for the microeconomics quiz, arguably a more specialized skill task, suggests that one should be cautious generalizing about effort responses to competition across task type.

2 Experimental Design

2.1 Microeconomics Quiz with No Selection into Participation

We administered the primary treatments to 2,415 students in 70 economics classes/sections at the University of California, Santa Barbara (UCSB) from fall 2013 through summer 2015. All sessions were held during the first two weeks of the relevant quarter. Before students entered the classroom sealed envelopes containing an entry survey and a microeconomics quiz were distributed across seats such that treatment groups were seated together. As students entered the class/section, they were randomly assigned to a treatment group by receiving a ticket from a shuffled deck that assigned them to a particular seat. Students who arrived late were asked to wait outside until the experiment ended. In all cases this was a very small number of students as we waited several minutes after the usual start time to close the door.

Once entry into the room had ceased, students completed an entry survey. This was a short

questionnaire asking about their age, race, gender, year in school, intended major, and the grade they expect to earn in the course in which the quiz was being administered (the survey is in Appendix A). Importantly, we classify students into two groups based on how they respond to the entry survey question, "What grade do you expect to earn in this class?". We define those who believe they are going to earn an "A" as high expectation students, and those who indicate they expect to earn a grade less than an "A" as low expectation students. This allows us to separately analyze effort responses to competition across perceived ability and tournament size, and perceived ability and payment structure. For context, 62% of students expect to earn an "A" in their current course. By linking the experiment data with student-level administrative files, we will explore differences between perceived ability and actual ability in Section 5.1.[8]

Next we explained that we were administering a microeconomics quiz in a large number of lower and upper division economics courses. We also informed participants that we would be coming around to explain how they could earn money for correctly answering quiz questions and that they would be entered into a $25 drawing at the end of the quiz to thank them for participating. At least in part because the quiz was administered at the beginning of classes/sections, participation was essentially 100 percent; fewer than five students left during the instruction phase and did not fill out a survey or participate in the quiz.

Students were given fifteen minutes to complete a ten question microeconomics quiz. They were not permitted to use a calculator or any other materials. We used six quiz forms to guard against cheating. The questions on each form were randomly drawn within subject category from the Test of Understanding in College Economics (TUCE) exam. The TUCE is a 30 question introductory microeconomics exam given to college economics students across the United States. An example of the quiz is in Appendix A. Students were incentivized for each correct quiz answer based on their randomly assigned treatment group. The treatment groups were as follows:

1. Piece Rate: Subjects earned $0.50 for each correct answer.

[8]While 62% students expect an "A" in their current course, only 18% of students actually earned an "A" in intermediate microeconomics (see Table 5).

2. Group of Two: The subject with the highest score in each pair earned $1 for each correct answer.

3. Group of Six, Winner Take All: The subject with the highest score in each group earned $3 for each correct answer.

4. Group of Ten, Winner Take All: The subject with the highest score in each group earned $5 for each correct answer.

5. Group of Six, Top Half Paid: The three highest scoring subjects in each group earned $1 for each correct answer.

6. Group of Ten, Top Half Paid: The five highest scoring subjects in each group earned $1 for each correct answer.

Following the foundational Gneezy et al. (2003) study, payments in all tournaments are a multiplicative version of the piece rate. Note that the ex ante expected value would be the same across all treatments if quiz scores were randomly assigned.

We implement several tournament sizes and payment structures to explore effort responses to both features for students of differing ability levels. The baseline treatment incentivizes students using a non-competitive piece rate payment scheme. Treatment two is a simple two person competition in which the winner is paid twice the piece rate. Treatment groups three and four are similar to two but for larger group sizes; they maintain a single winner. Together, treatments two, three and four – winner take all (WTA) treatments – allow us to compare treatments where the expected value of the tournament and proportion of winners remains constant but group size varies. One can think of this as scaling up a tournament; it is often referred to in the literature as organizational replication, see Orrison et al. (2004).

In contrast, treatments five and six maintain the proportion of winners – top half are paid (THP) – for larger group sizes. Comparing THP treatments with the WTA treatments allow us to hold group size and the expected value of the tournament constant, but vary the number of winners. Our

prior is that increasing the number of winners from a single winner (WTA) to the top half (THP) will increase effort of low types because now there is some positive probability, with a little luck, of earning a position in the top half of the group. In a winner take all payment structure, low types have no reason to exert costly effort as there is likely no chance they'll win. Alternatively, we expect high type effort to be unchanged or decreasing in the number of winners as they can still earn a spot in the top half even if they exert less effort. Of course, because our experiment takes place in a real life setting with a real-effort task, our experimental design does not perfectly map to the theory. One major difference is that there is a distribution of ability types in our setting rather than the two types modeled in Chen et al. (2011).

All subjects write a single exam under a single treatment; all comparisons are between-subject. Ties were broken by random draw. Given the number of groups, not all treatments were administered in every session, but more than one group was administered in every session. This is important because it allows for the inclusion of session (class/section) fixed effects.

All students were notified as to how and when they could see their score, and when and where winners could collect their payment within ten days of their session. All payments were made outside of the classroom where the student participated in the experiment.

2.2 Task Type

As it is possible that responses to competition depend on the type of task, in summer 2015 we also compared performance on the microeconomics quiz – a specialized skill task – to a relatively more general skill based quiz. This general skill quiz more closely resembles the tasks used in much of the previous experimental literature on competition; it required students to add and multiple two digit numbers without a calculator. To examine whether the effects of competition depend on task type, we randomly assigned 180 students in 11 sessions into one of four treatment groups: piece rate and a TUCE quiz, piece rate and an adding/multiplying quiz, group of two and a TUCE quiz, or group of two and an adding/multiplying quiz. For tractability, we limited group size to two in this round. The incentive structure is the same as above: $0.50 per correct answer in piece rate and

$1 for each correct answer for the highest scoring student in the group of two tournament. Because the adding/multiplying and TUCE quizzes were administered simultaneously in all 11 sessions, the former was formatted to look like a microeconomics quiz; calculating rent, minimum cost, average return, and so on. An example of the adding/multiplying quiz is included in Appendix A.

2.3 The Role of Selection into Participation

In contrast to most lab experiments, our environment features very different selection. In our setting, conditional on attending the chosen class/section, we have essentially 100 percent participation. This occurs by design because all experiments were run at the beginning of class/section. On the other hand, in a typical lab setting, students are recruited through a web based application and must show up to the lab at a designated time solely for the purpose of participating in the experiment. It has been shown that the sample of students who voluntarily participate in lab experiments may not be representative of the population from which they are drawn (Cleave et al., 2013).

In an effort to understand how our sample compares to a group that is relatively more selected, we ran 11 sessions in summer 2015 during the last 15 minutes of classes/sections. On these occasions, students could choose to stay and participate in the quiz or to leave. They were told that they could earn money for correctly answering quiz questions and that they would be eligible to win a $25 raffle for their participation. Roughly half of students in these sessions chose to stay and participate in the experiment.[9] Those who stayed were randomly assigned to a piece rate or group of two treatment. As in all other sessions, those in the piece rate group earned $0.50 for each correct answer and the top scorer in the group of two earned $1.00 for each correct answer. 137 students participated in these rounds.

[9]The type of selection induced by this design is discussed in detail in Section 5.3.

3 Subjects and Data

All data come from two sources. The main source is the data collected directly from the survey and experiment. The experiments were conducted in three waves: fall quarter 2013, spring quarter 2015 and summer 2015.[10] We ran experiments in 81 sessions with 2,732 total participants at UCSB. These data include quiz scores, treatment assignments, race, gender, major, academic year standing, age, and the grade they expect to earn in the course in which the quiz is administered. All tables are restricted to participants who responded to all questions on the entry survey. This restriction eliminates 52 participants. We also exclude data for 70 foreign exchange students as we do not have administrative data for these individuals. This leaves a final estimating sample of 2,610.[11] Table 1 reports summary statistics for quiz scores, standardized quiz scores, and expected grade disaggregated by the six main treatments (Panel A), by gender (Panel B), and class standing (Panel C). The sample in Table 1 is restricted to the 2,304 participants who took the TUCE quiz at the beginning of a class/section. We provide similar information for those who took the more general skill task and the rounds that were run at the end of classes/sections in Section 5. While column 1 reports raw TUCE quiz scores to give the reader some context, column 2, and all columns in all subsequent tables use scores that are standardized to mean zero and standard deviation one by course and quiz form.[12]

It is worth highlighting a few features of Table 1. The average TUCE quiz score is approximately 5 out of 10, the average male undergraduate outscores the average female undergraduate, and students in upper division courses outscore students in lower division courses. While the average score is also highest for the piece rate group, it is important to remember that there are no controls and not all treatments were assigned in every session. Approximately 62% of the participants have high expectations (expect to earn an "A" grade) in the current course and on average

[10] A potential concern is that subjects from the summer round are different from those in the school year rounds. We find no evidence of this; estimates from a balance test are reported in Appendix Table B1.

[11] All results are similar if these 122 observations are included and indicators for missing control variables are included as required.

[12] We standardize at this level to account for the fact that students become more proficient in economics as they progress through the major and that some of the quiz forms are more difficult than others.

women are somewhat less likely to have high expectations.

Panel A of Table 2 reports descriptive statistics for the pre-treatment characteristics collected on the entry survey for low expectation participants. Column 1 reports the mean and standard deviation for each characteristic for participants assigned to the group of two treatment. Columns 2-6 report the differences in mean characteristics between the group of two and each of the other treatment groups. Each entry in these columns comes from a separate regression.[13] There is little evidence of systematic differences in gender, race, major, class standing, or age across treatment groups. Panel B replicates this exercise for high expectation participants. Again, there is little evidence of imbalance across treatment groups, at least based on observable characteristics.

The experiment and survey data are also linked to administrative records that include the grades earned in core economics courses at UCSB. These data allow heterogeneity analysis that explores the relationship between expected grades and actual grades. These issues will be discussed in more detail in Section 5.1.

4 Empirical Specification

Our primary objective is to ask whether there are consistent patterns in participant effort when assigned to compete versus being assigned to a piece rate treatment that depends on perceived ability and tournament structure. More concretely: (1) Do low expectation participants change their effort level when assigned to compete in ways that are similar or different from high expectation participants? And, (2) Do these responses depend on tournament size or payment structure? In addition to affecting effort, we recognize that competition may lead to increased levels of stress which also may affect performance. While these are not observationally distinguishable, because the task and environment are relatively low stakes, it seems more likely that our estimated effects are capturing an effort response to competition. We therefore refer to it as effort, while acknowledging that other factors may be absorbed as well. We use the following simple specification to examine the above

[13] We report balance by comparing each group to the group of two because this is the treatment that was administered in all sessions.

questions.

$$Y_{iags} = \alpha + \theta_{ag} + \gamma_s + X_{iags}\beta + \varepsilon_{iags} \quad (1)$$

Y_{iags} is the standardized quiz score for student i, of perceived ability a (low and high expectation students), in treatment group g (piece rate, group of two, large group WTA, large group THP), in session s. For descriptive ease, we collapse the two larger winner take all tournaments (group of six WTA and group of ten WTA) into one group which we call large group WTA. Similarly, we construct large group THP by collapsing the two larger top half paid treatments (group of six THP and group of ten THP).[14] As such, θ is a vector of seven indicator variables for expected grade specific treatment groups. Low expectation students assigned to the piece rate treatment is the omitted category. γ_s is a vector of session fixed effects. Note that session indicators absorb time because classes/sections are quarter and academic year specific. X is a vector of student background characteristics, including race, gender, major, and year in school. Because treatment is randomly assigned within sessions, all results are similar regardless of which controls are included. ε is the usual error term and all standard errors are clustered at the session level.

5 Results

5.1 Economics Quiz at the Beginning of Section

The results for equation 1 are reported in Table 3. Unless otherwise specified, all point estimates in each table come from a single regression. Column 1 reports the difference in quiz score between the piece rate treatment and each of the three tournament groups for low expectation participants. P-values for the difference in competition effect between the specified larger tournament size and the group of two for low expectation students are reported in square brackets under the standard errors. Column 2 reports the same set of results for high expectation students. In other words, each

[14] Collapsing in this way has no qualitative impact. The results disaggregating the size six and ten groups are reported in Appendix B.

entry is the average difference in score between the specified group and the piece rate group for individuals who expect a high grade. Similarly, the p-values reported in square brackets are for the difference in the competition effect between the group of two and the specified larger group for high expectation students. Column 3 reports the difference between the low and high expectation student groups within each tournament type.

Column 1 reveals that low expectation students reduce their effort when forced to compete. More specifically, the average score in the tournaments is 16.2-32.7% of a standard deviation lower than the average piece rate score; for context, 25% of a standard deviation is approximately half a point on the ten-point TUCE quiz.[15] Further, there is at best limited evidence that the size of the tournament or the payment structure matter; while we reject the null hypothesis that the point estimate for large groups under WTA is the same as for a group of two, at the finer gradation reported in Table B4, we see no such pattern. In contrast, there are no precisely estimated competition effects for high expectation students at any tournament size. Column 3 shows that the competition effects for the two ability groups are statistically different from one another.[16]

These results naturally lead one to wonder whether expected grades reflect actual ability/skill or over/under confidence of some type for some subgroups. Our administrative data on grades in core economics courses allow us to probe this issue. Table 4 interacts expected grade with earned grade, where earned grade is defined as the participant's first intermediate microeconomics grade. If no intermediate theory grade is available, we use their principles grade. The sample size is slightly smaller for this table because we do not have intermediate or principles grades for a small number of junior college transfer students. We divide students into two groups: those who earn less than an "A" (non-A students) and those who earn an "A" (A students).

Not surprisingly, as we sub-sample or interact to greater degrees, many estimates become quite noisy. What is clear, is that the subset of subjects who are non-A students who also have low

[15] As a way to check for "giving-up" behavior, we examined average scores on the first and second halves of the quiz. We found no evidence of such behavior.

[16] Appendix Tables B2-B4 report results for the disaggregated set of treatments including all four of the larger groups (group of 6 WTA, group of 6 THP, group of 10 WTA and group of 10 THP). Results are similar across all specifications.

expectations put forth less effort when assigned to competitive treatments (column 1), and this competition effect is statistically different from the competition effect for high expectation non-A students (compare columns 1 and 3). In summary, we can rule out large negative effects for non-A students with high expectation suggesting that optimism or overconfidence shields this group from the negative competition effect. And, while the other comparisons are noisy (i.e, comparing columns 1 and 2), what we can say with some confidence is that reduced effort when confronted with competition seems to be driven by weaker students who realize that they are weaker students, a result that is more consistent with honest self-reflection than under confidence.

At least part of the reason that the low expectation non-A student group is driving the results is because this is where the discrepancy between earned and expected grades exists. Table 5 shows that among A students, about 81% expect to earn an "A" in the current course. On the other hand, among non-A students, 42% expect less than an "A" and 58% expect an "A".[17]

Many previous studies identify important differences in effort response to competition by gender. In a variety of field and lab settings, it has been shown that men tend to increase their effort when forced to compete, while women's performance is unchanged (Gneezy et al., 2003; Gneezy and Rustichini, 2004). Appendix Table B5 replicates Table 4 but instead of earned grade, we interact expected grade with gender. Columns 1 and 2 (3 and 4) report the competition effect for low (high) expectation men and women. In contrast to previous studies, in our setting the negative competition effect point estimate is most negative for low expectation men, but due to imprecision, we cannot reject that the male and female effects are the same for students within expectation groups. Overall, these findings are consistent, but not definitive, evidence that the competition effects are driven by less able men exerting less effort when forced to compete.[18]

[17] Replicating this for men and women separately reveals few differences.

[18] We also find no evidence of gender differences in the response to the sex composition of the group.

5.2 Task Type

In contrast to many previous studies, in our setting we find that competition reduces effort, at least for an important subset of participants. In this section we investigate two possible features of our experiment that might explain the divergence of our results from past findings that show performance improves, at least among some sub-groups, when subjects are forced to compete. First, we explore the potential role of task type. In our design, subjects take a microeconomics quiz. Relative to the tasks that are often implemented, such as solving mazes or adding up two-digit numbers, the task in this experiment draws on a more specialized skill. Second, our sample is selected on major or course enrollment (largely economics and accounting students), but has almost no selection on the showing up or signing up margin since all quizzes took place at the start of classes/sections and essentially 100 percent of students attending class/section participated. We examine each of these in turn.

To test whether a specialized skill task in a competitive environment has a different effect on performance than a more general skill task, we implement an additional treatment. In this treatment students were asked to multiply and add sequences of numbers without a calculator. These simple calculations were formatted to look like microeconomics questions because this treatment was administered in the same session as the microeconomics quiz task. For example, questions from the more general skill task included calculating rent, minimum cost, and average return. For simplicity, we only included one competitive incentive scheme, group of two (we call this compete), and the piece rate treatment. Table 6 reports summary statistics for these additional rounds. For comparative purposes, column 1 reports the average raw quiz score, the average standardized score, the percent of the sample with high expectations, and the percent female for participants assigned to the piece rate and group of two standard microeconomics quiz treatments. Column 2 reports the same summary measures for participants assigned to the general skill task, again including both piece rate and groups of two. The estimating sample is 172 students.[19] It is important

[19] While 180 students participated in these rounds, 8 were dropped due to the sample restrictions outlined in Section 3.

to note that you cannot compare the average raw scores across columns 1 and 2 because the tests are very different.

We use an empirical specification similar to equation 1. The primary differences are as follows. First, there are only two compensation schemes: group of two and piece rate (piece rate for low expectation students continues to be the omitted category). Second, the model now includes an indicator variable for the general skill task and this indicator is interacted with the three treatment group indicators (low expectation group of two, high expectation piece rate, and high expectation group of two). Column 1 of Table 7 reports the results. As in the main results reported in Table 3, low expectation participants exert less effort on the microeconomics quiz when forced to compete, while the effort of high expectation students does not change (see Panel B). Low expectation students score 18.4% of a standard deviation lower in the group of two treatment compared to the piece rate treatment. In contrast, low expectation students assigned to the adding/multiplying task score higher when assigned to the competitive treatment, although the point estimate is noisy. The difference-in-difference estimate shows that the competition effect for the general skill task is 40.0% of a standard deviation higher than the competition effect for the specialized skill task. Even if the competition effect for the general skill task is indistinguishable from zero, we can reject that it is as negative as the competition effect for the specialized skill task for the low expectation group. Taken together, these results suggest that effort in competitive environments depends importantly on the nature of the task. And if the task involves a specialized skill, it also depends on ability. Unsurprising, our general skill treatment yields results similar to what has previously been found in the experimental literature, as those studies implement tasks involving similar skills. Results for the microeconomics quiz task are quite different.

5.3 Selection into Participation

In order to explore the potential for at least some forms of selection to impact the results, we ran 11 additional sessions in which we allowed more selection into participation. In contrast to all other sessions, these sessions were run at the end of class/section. Students were invited to stay and

take the microeconomics quiz under the same set of incentives as before. We are, of course, not suggesting that all forms of selection must be identical. The form of selection we induce is caused by voluntary staying to participate, while lab experiments induce selection based on voluntary signing up and showing up to participate. The non-randomness of the subjects could certainly be different across these two cases.

On average, approximately half of the students present chose to participate (compared to essentially 100 percent in the beginning of class/section sessions). Summary statistics for these additional end of class/section sessions are reported in last column of Table 6. There are 134 students in the estimating sample.[20] There is some evidence of selection; stayers score slightly better on average and are somewhat more likely to be male. Of course, our primary question is whether this selection implies a different response for those who are randomly assigned to compete.

The results are reported in column 2 of Table 7. The specification is identical to that exploring the difference between the microeconomics quiz task and the general skill quiz in column 1, except that the indicator for the new treatment (the end of class/section treatment) cannot be identified separately from session fixed effects. We can, however, identify this indicator interacted with the three included treatment group indicators (low expectation group of two, high expectation piece rate, and high expectation group of two). In contrast to the differential response to competition in specialized and general skill tasks, the response to competition in sessions run at the beginning of class/section and at the end of class/section (less and more selected participation cases) are both negative for low expectation students, and we cannot reject the null hypothesis that the point estimates are the same. And as in all other cases, there is no evidence of an effort response for high expectation students. Our end of session results suggest that the reason our main findings differ from previous studies is unlikely to be the result of selection, unless lab participation induces a very different and very strong form of non-randomness than the offer to exit the experiment.

[20] While 137 participated, 3 are dropped due to the sampling restrictions outlined in Section 3.

6 Conclusion

In a classroom setting where students are asked to complete a real-effort task, we show that individuals who believe they are lower scoring students reduce their effort when forced to compete, while there are few detectable effects for students who expect to earn higher grades. We find that students who believe they will earn less than an "A" grade in their current economics course score about 16-33% of a standard deviation lower when forced to compete compared to the piece rate treatment, and that this effect is driven by men and low expectation students who have realistic beliefs about their ability. In contrast, previous research has tended to find that men increase their effort level when forced to compete. Perhaps the biggest take away from this experiment is the evidence that our results diverge from past findings because of task type differences and not because of selection. The evidence clearly suggests that task type plays a critical role in one's response to competition. Stated more generally, responses to competition are likely to depend importantly on the environment and the individual's position in the skill distribution; people who might increase their effort when faced with competition in one environment might decrease or not change their effort in another environment. As such, one should be cautious, perhaps even skeptical, about generalizing results about the distribution of responses to competition from one context to another.

References

Andreoni, James and Andy Brownback, "All Pay Acutions and Group Size: Grading on a Curve and Other Applications," *Journal of Economic Behavior & Organization*, 2017, *137*, 361–373.

Angrist, Joshua and Victor Lavy, "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial," *American Economic Review*, 2009, *99* (4), 1384–1414.

_ **, Daniel Lang, and Philip Oreopoulos**, "Incentives and Services for College Achievement: Evidence from a Randomized Trial," *American Economic Journal: Applied Economics*, 2009, *1* (1), 136–63.

Barut, Yasar, Dan Kovenock, and Charles N Noussair, "A Comparison of Multiple-Unit All-Pay and Winner-Pay Auctions Under Incomplete Information," *International Economic Review*, 2002, *43* (3), 675–708.

Becker, William E and Sherwin Rosen, "The Learning Effect of Assessment and Evaluation in High School," *Economics of Education Review*, 1992, *11* (2), 107–118.

Betts, Julian R and Jeff Grogger, "The Impact of Grading Standards on Student Achievement, Educational Attainment, and Entry-Level Earnings," *Economics of Education Review*, 2003, *22* (4), 343–352.

Booth, Alison and Patrick Nolen, "Choosing to Compete: How Different are Girls and Boys?," *Journal of Economic Behavior & Organization*, 2012, *81* (2), 542–555.

Brown, Jennifer, "Quitters Never Win: The (Adverse) Incentive Effects of Competing with Superstars," *Journal of Political Economy*, 2011, *119* (5), 982–1013.

Brownback, Andy, "A Classroom Experiment on Effort Allocation Under Relative Grading," *Economics of Education Review*, 2018, *62*, 113–128.

Bull, Clive, Andrew Schotter, and Keith Weigelt, "Tournaments and Piece Rates: An Experimental Study," *Journal of Political Economy*, 1987, *95* (1), 1–33.

Burgess, Simon, Robert Metcalfe, and Sally Sadoff, "Understanding the Response to Financial and Non-Financial Incentives in Education: Field Experimental Evidence Using High-Stakes Assessments," 2016. Working Paper.

Cason, Timothy N, William A Masters, and Roman M Sheremeta, "Entry into Winner-Take-All and Proportional-Prize Contests: An Experimental Study," *Journal of Public Economics*, 2010, *94* (9-10), 604–611.

Chen, Hua, Sung H Ham, and Noah Lim, "Designing Multiperson Tournaments with Asymmetric Contestants: An Experimental Study," *Management Science*, 2011, *57* (5), 864–883.

Cleave, Blair L, Nikos Nikiforakis, and Robert Slonim, "Is There Selection Bias in Laboratory Experiments? The Case of Social and Risk Preferences," *Experimental Economics*, 2013, *16* (3), 372–382.

Dechenaux, Emmanuel, Dan Kovenock, and Roman M Sheremeta, "A Survey of Experimental Research on Contests, All-Pay Auctions and Tournaments," *Experimental Economics*, 2015, *18* (4), 609–669.

Dreber, Anna, Emma von Essen, and Eva Ranehill, "Gender and Competition in Adolescence: Task Matters," *Experimental Economics*, 2014, *17* (1), 154–172.

Figlio, David N and Maurice E Lucas, "Do High Grading Standards Affect Student Performance?," *Journal of Public Economics*, 2004, *88* (9-10), 1815–1834.

Fischer, Stefanie, "The Downside of Good Peers: How Classroom Composition Differentially Affects Men's and Women's STEM Persistence," *Labour Economics*, 2017, *46*, 211–226.

Garratt, Rodney J, Catherine Weinberger, and Nick Johnson, "The State Street Mile: Age and Gender Differences in Competition Aversion in the Field," *Economic Inquiry*, 2013, *51* (1), 806–815.

Gneezy, Uri and Aldo Rustichini, "Gender and Competition at a Young Age," *American Economic Review Papers and Proceedings*, 2004, *94*, 377–381.

_ , Kenneth L Leonard, and John A List, "Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society," *Econometrica*, 2009, *77*, 1637–1664.

_ , Muriel Niederle, and Aldo Rustichini, "Performance in Competitive Environments: Gender Differences," *Quarterly Journal of Economics*, 2003, *118* (3), 1049–1074.

Green, Jerry R and Nancy L Stokey, "A Comparison of Tournaments and Contracts," *Journal of Political Economy*, 1983, *91*, 349–364.

Günther, Christina, Neslihan Arslan Ekinci, Christiane Schwieren, and Martin Strobel, "Women Can't Jump? - An experiment on competitive attitudes and stereotype threat," *Journal of Economic Behavior & Organization*, 2010, *75* (3), 395–401.

Gupta, Nabanita Datta, Anders Poulsen, and Marie Claire Villeval, "Gender Matching and Competitiveness: Experimental Evidence," *Economic Inquiry*, 2013, *51* (1), 816–835.

Harbring, Christine and Bernd Irlenbusch, "An Experimental Study on Tournament Design," *Labour Economics*, 2003, *10* (4), 443–464.

Jalava, Nina, Juanna Schrøter Joensen, and Elin Pellas, "Grades and Rank: Impacts of Non-Financial Incentives on Test Performance," *Journal of Economic Behavior & Organization*, 2015, *115*, 161–196.

Jr, Roland G Fryer, "Financial Incentives and Student Achievement: Evidence from Randomized Trials," *The Quarterly Journal of Economics*, 2011, *126* (4), 1755–1798.

Kremer, Michael, Edward Miguel, and Rebecca Thornton, "Incentives to Learn," *American Economic Journal: Applied Economics*, 2009, *99* (3), 437–456.

Lazear, Edward P and Sherwin Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," 1981.

Leuven, Edwin, Hessel Oosterbeek, and Bas Van der Klaauw, "The Effect of Financial Rewards on Students' Achievement: Evidence from a Randomized Experiment," *Journal of the European Economic Association*, 2010, *8* (6), 1243–1265.

Lim, Wooyoung, Alexander Matros, and Theodore Turocy, "Bounded Rationality and Group Size in Tullock Contests: Experimental Evidence," *Journal of Economic Behavior & Organization*, 2014, *99*, 155–167.

List, John, Daan Van Soest, Jan Stoop, and Haiwen Zhou, "On the Role of Group Size in Tournaments: Theory and Evidence from Lab and Field Experiments," Technical Report, National Bureau of Economic Research 2014.

Müller, Wieland and Andrew Schotter, "Workaholics and Dropouts in Organizations," *Journal of the European Economic Association*, 2010, *8* (4), 717–743.

Niederle, Muriel and Lise Vesterlund, "Do Women Shy Away from Competition? Do Men Compete Too Much?," *Quarterly Journal of Economics*, 2007, *122*, 1067–1101.

Oettinger, Gerald S, "The Effect of Nonlinear Incentives on Performance: Evidence from "Econ 101"," *The Review of Economics and Statistics*, 2002, *84* (3), 509–517.

Orrison, Alannah, Andrew Schotter, and Keith Weigelt, "Multiperson Tournaments: An Experimental Examination," *Management Science*, 2004, *50* (2), 268–279.

Prendergast, Canice, "The Provision of Incentives in Firms," *Journal of Economic Literature*, 1999, *37* (1), 7–63.

Sutter, Matthias and Daniela Glätzle-Rützler, "Gender Differences in the Willingness to Compete Emerge Early in Life and Persist," *Management Science*, 2014, *61* (10), 2339–2354.

Villeval, Marie Claire, "Ready, Steady, Compete," *Science*, 2012, *335* (3), 544–545.

Table 1: Summary Statistics - Test Scores (TUCE) and Expected Grades

	Score	Standardized Score	E(A)	Sample Size
	(1)	(2)	(3)	(4)
Panel A: Treatment Group				
Piece rate	5.50	-0.02	0.57	422
	(2.04)	(1.00)	(0.50)	
Group of Two	5.14	0.02	0.60	642
	(2.06)	(1.00)	(0.49)	
Large Group WTA	4.96	-0.02	0.64	610
	(1.99)	(0.99)	(0.48)	
Large Group THP	4.95	0.02	0.65	630
	(1.96)	(0.97)	(0.48)	
Panel B: Gender				
Male	5.36	0.12	0.64	1,381
	(2.02)	(1.00)	(0.48)	
Female	4.73	-0.17	0.58	923
	(1.96)	(0.95)	(0.49)	
Panel C: Class Standing				
Lower Division	4.84	0.00	0.61	1,678
	(1.97)	(0.99)	(0.49)	
Upper Division	5.82	0.03	0.64	626
	(1.99)	(0.99)	(0.49)	

Notes: Mean scores and expected grades are reported by subgroups. Standard deviations are in parentheses. E(A) denotes students who expect to earn an A (high expectation) and E(NA) denotes students who do not expect to earn an A (low expectation). WTA are treatments where the top scorer is paid, THP are treatments where the top half of the group is paid.

Table 2: Pre-Treatment Characteristics - Sample Balance

		Differences Between Treatments		
	Group of 2 Mean	Piece Rate	Large Group WTA	Large Group THP
	(1)	(2)	(3)	(4)
Panel A: E(NA)				
Female	0.41	-0.04	0.08	0.15
	(0.49)	[0.49]	[0.19]	[0.01]
Hispanic	0.18	0.06	-0.03	-0.02
	(0.39)	[0.23]	[0.56]	[0.71]
Black	0.02	-0.01	0.02	0.02
	(0.12)	[0.60]	[0.35]	[0.16]
Asian	0.37	-0.09	-0.03	-0.01
	(0.48)	[0.18]	[0.54]	[0.90]
Other	0.07	-0.01	-0.04	-0.04
	(0.26)	[0.70]	[0.21]	[0.17]
Accounting Major	0.42	0.09	-0.09	-0.01
	(0.49)	[0.14]	[0.08]	[0.82]
Other Major	0.30	-0.09	0.08	0.06
	(0.46)	[0.07]	[0.06]	[0.18]
Sophomore	0.30	0.11	0.05	0.00
	(0.46)	[0.03]	[0.33]	[0.93]
Junior	0.36	-0.07	0.05	0.07
	(0.48)	[0.22]	[0.28]	[0.15]
Senior	0.19	0.00	-0.07	-0.07
	(0.39)	[0.95]	[0.03]	[0.01]
Age 20-21	0.38	0.05	0.02	0.04
	(0.49)	[0.45]	[0.72]	[0.37]
Age 22+	0.17	-0.04	0.03	-0.06
	(0.37)	[0.40]	[0.47]	[0.11]
Sample Size	256	439	477	478
Panel B: E(A)				
Female	0.39	-0.09	-0.04	0.02
	(0.49)	[0.10]	[0.38]	[0.56]
Hispanic	0.14	-0.03	-0.03	0.01
	(0.35)	[0.41]	[0.38]	[0.78]
Black	0.02	0.00	0.00	0.00
	(0.13)	[0.88]	[0.87]	[0.67]
Asian	0.45	0.01	0.02	0.02
	(0.50)	[0.86]	[0.63]	[0.70]
Other	0.05	-0.03	0.02	0.01
	(0.21)	[0.20]	[0.41]	[0.54]
Accounting Major	0.42	-0.01	0.05	-0.04
	(0.50)	[0.84]	[0.15]	[0.29]
Other Major	0.29	0.08	-0.03	-0.01
	(0.46)	[0.05]	[0.25]	[0.81]
Sophomore	0.27	-0.05	-0.02	-0.01
	(0.45)	[0.24]	[0.57]	[0.80]
Junior	0.39	0.04	0.01	-0.02
	(0.49)	[0.43]	[0.74]	[0.31]
Senior	0.19	0.03	-0.03	-0.01
	(0.39)	[0.34]	[0.31]	[0.80]
Age 20-21	.39	0.06	0.00	0.00
	(0.49)	[0.29]	[0.93]	[0.91]
Age 22+	0.22	0.01	-0.05	-0.05
	(0.42)	[0.84]	[0.12]	[0.09]
Sample Size	386	625	775	794

Notes: Column 1 reports means for the group of two treatment. Corresponding standard deviations are in parentheses. Differences in columns 2-6 are from OLS regressions that include session fixed effects. Sample sizes listed in columns 2-6 include the group of two and the group listed at the top of each column. P-values for differences are in square brackets. E(A) denotes students who expect to earn an A (high expectation) and E(NA) denotes students who do not expect to earn an A (low expectation). WTA are treatments where the top scorer is paid, THP are treatments where the top half of the group is paid.

Table 3: Score Differences by Incentive Structure Across Grade Expectation

	Difference Estimates		Effects Statistically Different?
	E(NA)	E(A)	(1)-(2)
	(1)	(2)	(3)
Group of Two - Piece Rate Difference	-0.162*	0.059	-0.221**
	(0.087)	(0.088)	(0.107)
Large Group WTA - Piece Rate Difference	-0.327***	-0.038	-0.289**
	(0.122)	(0.115)	(0.121)
	[0.06]	[0.25]	
Large Group THP - Piece Rate Difference	-0.264**	0.012	-0.276**
	(0.117)	(0.121)	(0.138)
	[0.23]	[0.60]	
Equivalence of Group Coefficients [P-Values]	0.15	0.50	

Notes: Sample size is 2,304. All coefficients are from a single regression (equation 1) that includes indicators for race, gender, major, and year in school, as well as session fixed effects. E(A) denotes students who expect to earn an A (high expectation) and E(NA) denotes students who do not expect to earn an A (low expectation). WTA are treatments where the top scorer is paid and THP are treatments where the top half of the group is paid. The 6 and 10 person WTA (6 and 10 person THP) tournament treatments are collapsed into a single group category denoted as Large Group WTA (Large Group THP). Standard errors are clustered at the session level and are reported in parentheses. Square brackets report the p-value testing for the equivalence of the corresponding coefficient and the coefficient for the group of two. ** (*) indicates statistical significance at the 5 (10) percent level.

Table 4: Score Differences by Incentive Structure Across Grade Expectations by Previous Microeconomics Grade

	E(NA)		E(A)	
	Non-A Student	A Student	Non-A Student	A Student
	(1)	(2)	(3)	(4)
Group of Two - Piece Rate Difference	-0.155*	-0.070	0.143	-0.177
	(0.088)	(0.278)	(0.101)	(0.203)
Large Group WTA - Piece Rate Difference	-0.291**	-0.338	-0.008	-0.002
	(0.120)	(0.397)	(0.123)	(0.204)
	[0.13]	[0.38]	[0.16]	[0.20]
Large Group THP - Piece Rate Difference	-0.297**	0.102	0.100	-0.162
	(0.116)	(0.311)	(0.120)	(0.196)
	[0.10]	[0.46]	[0.65]	[0.92]
Equivalence of Group Coefficients [P-Values]	0.18	0.52	0.28	0.35

Notes: Sample size is 2,258. Non-A students earned less than an A and A students earned an A in microeconomics at UCSB. All coefficients are from a single regression (equation 1) that includes indicators for race, gender, major, and year in school, as well as session fixed effects. E(A) denotes students who expect to earn an A (high expectation) and E(NA) denotes students who do not expect to earn an A (low expectation). WTA are treatments where the top scorer is paid and THP are treatments where the top half of the group is paid. The 6 and 10 person WTA (6 and 10 person THP) tournament treatments are collapsed into a single group category denoted Large Group WTA (Large Group THP). Standard errors are clustered at the session level and are reported in parentheses. ** (*) indicates statistical significance at the 5 (10) percent level.

Table 5: Earned Microeconomics Grade and Current Expected Grade

	Earned Intermediate Microeconomics Grade	
	Non-A Student	A Student
E(Grade) in Current Class		
E(NA)	785 (42%)	77 (19%)
E(A)	1,066 (58%)	330 (81%)

Notes: E(Grade) is the grade students expect to earn in the current course. E(A) denotes students who expect to earn an A (high expectation) and E(NA) denotes students who do not expect to earn an A (low expectation). Non-A students earned less than an A and A students earned an A in intermediate microeconomics at UCSB. More precisely, the earned grade is defined as the participant's first intermediate microeconomics grade. If no intermediate theory grade is available, we use their principles grade. Each cell is the number of students within a column who report a particular belief about their grade. The corresponding share is in parentheses.

Table 6: Summary Statistics for Mechanism Exploration Sessions

	Start of Section Task=TUCE (1)	Start of Class/Section Task=Adding (2)	End of Class/Section Task=TUCE (3)
Score	5.28	7.20	5.99
	(2.06)	(2.05)	(1.91)
Standardized Score	0.01	0.00	0.23
	(1.00)	(1.01)	(0.92)
E(A)	0.59	0.56	0.60
	(0.49)	(0.50)	(0.49)
Female	0.40	0.46	0.35
	(0.49)	(0.50)	(0.48)
Upper Division	0.37	0.44	0.54
	(0.48)	(0.50)	(0.50)
Sample Size	1,064	172	134

Notes: Means are reported by subgroup. Standard deviations are in parentheses. Start of Class/Section and End of Class/Section refer to rounds of the experiment run at the beginning and end of the classes/sections. Adding refers to the general skill task and TUCE to the introductory microeconomics quiz, the ability-specific task. E(A) denotes students who expect to earn an A (high expectation).

Table 7: Score Differences: TUCE Task Compared to Adding Task and Beginning of Class/Section TUCE Compared to End of Class/Section TUCE

	Start of Class/Section Task=Adding or TUCE (1)	End or Start of Class/Section Task=TUCE (2)
Panel A: Coefficients		
E(NA) Group of Two	-0.184**	-0.187**
	(0.090)	(0.091)
E(A) Piece Rate	0.006	0.000
	(0.096)	(0.096)
E(A) Group of Two	0.062	0.064
	(0.102)	(0.102)
Comparison Treatment (CT)†	-0.109	–
	(0.193)	
CT E(NA) Group of Two	0.400*	0.047
	(0.215)	(0.271)
CT E(A) Piece Rate	0.428*	0.169
	(0.222)	(0.304)
CT E(A) Group of Two	0.333	0.197
	(0.231)	(0.243)
Panel B: Relevant D-in-D Estimates		
Competition Effect for E(NA) TUCE	-0.184**	-0.187**
	(0.090)	(0.091)
Competition Effect for E(NA) CT	0.213	-0.141
	(0.202)	(0.254)
D-in-D	0.400*	0.047
	(0.215)	(0.271)
Competition Effect for E(A) TUCE	0.056	0.065
	(0.091)	(0.091)
Competition Effect for E(A) CT	-0.038	0.092
	(0.155)	(0.260)
D-in-D	-0.094	0.028
	(0.174)	(0.274)
Sample Size	1,236	1,198

Notes: Each column is from a single regression. All models include indicators for race, gender, major, and year in school, as well as session fixed effects. Start of Class/Section and End of Class/Section refer to rounds of the experiment run at the beginning and end of the classes/sections. Adding refers to the general skill task and TUCE to the introductory microeconomics quiz, the ability-specific task. E(A) denotes students who expect to earn an A (high expectation) and E(NA) denotes students who do not expect to earn an A (low expectation). Standard errors are clustered at the session level and are reported in parentheses. ** (*) indicates statistically significance at the 5 (10) percent level.

† In column 1 the comparison group (CT) indicates the adding task and in column 2 it indicates the end of class/section treatment.

Appendix A

Entry Survey

You are being asked to participate in a study by Kelly Bedard, Stefanie Fischer, and Jon Sonstelie. You may only participate once. For your participation today, we will enter you in a lottery in which one person in this class will receive $25 cash today (photo ID required). If you are younger than 18 you are not eligible for the lottery. While those under 18 years of age can participate in the tournament, your data will not be used for research purposes.

You have also been selected to receive the opportunity to compete against the person with the same lettered quiz sitting near you. The highest scoring person in your pair wins $1 for each of their correct answers. The microeconomics quiz includes 10 randomly selected questions. You have 15 minutes to complete the quiz. Ties will be broken by a random draw.

Your score and your payment status will be available in approximately one week. All payments will be made outside the classroom at the beginning or end of class.

We are conducting a study to assess proficiency in foundational microeconomics and analyze competition and test taking. By signing up for this experiment, you are acknowledging that the authors of this study will follow your Economics academic records at UCSB from the beginning of your enrollment through summer 2016. There is minimal foreseeable risk associated with participation; some participants may find some questions frustrating. On the other hand, the potential benefits to UCSB and society associated with a better understanding of the factors that improve test performance are potentially large. This data will not be used for any other purpose nor will any identifiable information ever be made public. All identifying data will be held in confidence from all instructors until after this academic quarter. That being said, absolute confidentiality cannot be guaranteed, since research documents are not protected from subpoena.

Your participation is voluntary. There will be no repercussions should you decide not to participate. You may withdraw your participation at any time and remain eligible for the lottery. If you have questions you may contact Kelly Bedard at (805) 893-5571 or kelly@econ.ucsb.edu. If you have any questions regarding your rights and participation as a research subject, please contact the Human Subjects Committee at (805) 893-3807 or hsc@research.ucsb.edu. Or write to the University of California, Human Subjects Committee, Office of Research, Santa Barbara, CA 93106-2050.

By signing below, you acknowledge the above information.

<u>We would like to ask you a few questions</u> (please bubble in answers on scantron)

1.	What is your sex?	(A) Female	(B) Male			
2.	How old are you?	(A) 17	(B) 18 or 19	(C) 20 or 21	(D) 22 or 23	(E) 24+
3.	Are you Hispanic/Latino?	(A) Yes	(B) No			
4.	What is your race?	(A) White	(B) Black	(C) Asian	(D) Other	
5.	Academic Year?	(A) Freshman	(B) Sophomore	(C) Junior	(D) Senior	
6.	Major/Intended Major?	A) Economics	(B) Economics & Accounting	(C) Economics & Mathematics	(D) Other	(E) Undecided
7.	What grade do you expect in this course?		(A) D+ or lower	(B) C-, C or C+	(C) B-, B or B+	(D) A- or A

_____ _____

Print name Signature

_____ _____

Date Perm #

_____ _____

Primary e-mail address Local phone number

Form 1 of the TUCE Quiz (Specialized Skill Task)

8. Suppose a city facing a shortage of rental apartments eliminates rent controls. Which of the following is most likely to occur?

   A. a decrease in rents and a decrease in the number of apartment units supplied
   B. an increase in rents and an increase in the number of apartment units supplied
   C. a decrease in the demand for apartments and an increase in the number of apartment units supplied
   D. an increase in the demand for apartments and a decrease in the number of apartment units supplied

9. If all of the firms in a competitive industry are legally required to meet new regulations that increase their cost of production:

   A. supply of the product will decrease.
   B. demand for the product will decrease.
   C. the long-run economic profits of individual firms in the industry will decrease.
   D. the short-run economic profits of individual firms in the industry will increase.

10. At the profit-maximizing level of output, a perfectly competitive firm will:

    A. produce the quantity of output at which marginal cost equals price.
    B. produce the quantity of output at which marginal cost is minimized.
    C. keep marginal cost lower than price, so profits will be greater than zero.
    D. try to sell all the output it can produce, to spread fixed costs across the largest possible number of units.

11. A state legislature increased the tax on gasoline sold in the state from $.20 to $.30 per gallon. A supporter said the tax would "make the distribution of after-tax income in the state more equal." This statement would be true only if it could be shown that, after the tax is increased:

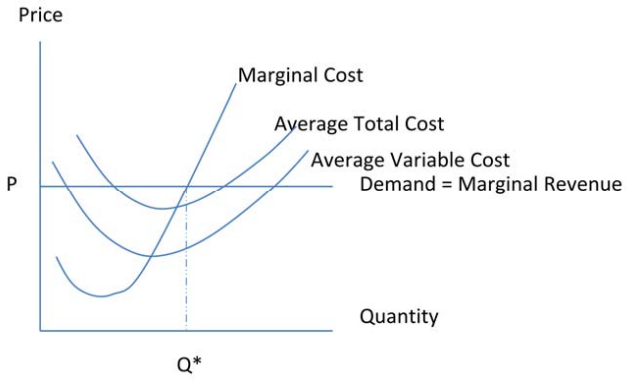
    A. people with low incomes buy more gasoline than people with high incomes.
    B. the quantity of gasoline purchased in the state is highly responsive to changes in price.
    C. people with high incomes tend to spend the same proportion of their incomes on gasoline as people with low incomes.
    D. people with high incomes tend to spend a larger proportion of their incomes on gasoline than people with low income.

12. The opportunity cost of being a full-time student at a university instead of working full-time at a job includes all of the following EXCEPT:

    A. payments for meals.
    B. payments for tuition.
    C. payments for books.
    D. income from the full-time job.

13. "Water is essential to life, but inexpensive to buy." Which of the following best explains this observation?

A. Water has a high total utility, but a low marginal utility.
B. Water has a low total utility, but a high marginal utility.
C. The quantity supplied of water is less than the quantity demanded at the market price.
D. The quantity supplied of water is greater than the quantity demanded at the market price.



14. Which of the following is true for this profit-maximizing firm at price P in the graph above?

A. It is not earning any economic profits.
B. It is currently earning short-run economic profits.
C. It should shut down to minimize its economic losses.
D. It will continue to earn economic profits in the long run.

15. If the exchange rate between dollars ($) and yen (¥) changes from $1 = ¥200 to $1 = ¥100, and domestic prices in both countries stay the same, has the dollar appreciated or depreciated, and would U.S. imports from Japan become less expensive or more expensive?

	Value of the dollar	U.S. imports from Japan
A.	Appreciated	Less expensive
B.	Appreciated	More expensive
C.	Depreciated	Less expensive
D.	Depreciated	More expensive

16. Suppose the only two cola companies (Acola and Bcola) in a nation are deciding whether to charge high or low prices for their colas. The companies' price strategies are shown in the table below. The four pairs of payoff values show what each company expects to earn or lose in millions of dollars, depending on what the other company does.

		Acola's Price Strategy	
		High Price	Low Price
Bcola's Price Strategy	High Price	Acola +$100 Bcola +$100	Acola +$250 Bcola -$50
	Low Price	Acola -$50 Bcola +$250	Acola +$50 Bcola +$50

If both companies believe that most consumers are soon going to quit drinking colas, and switch to other products, what is the equilibrium outcome?

A. Both Acola and Bcola will charge a low price.
B. Both Acola and Bcola will charge a high price.
C. Acola will charge a low price; Bcola will charge a high price.
D. Acola will charge a high price; Bcola will charge a low price.

17. In Sunshine City, one local ice cream company operates in a competitive labor market and product market. It can hire workers for $45 a day and sell ice cream cones for $1.00 each. The table below shows the relationship between the number of workers hired and the number of ice cream cones produced and sold.

Number of Workers Hired	Number of Ice Cream Cones Sold
4	340
5	400
6	450
7	490
8	520

As long as the company stays in business, how many workers will it hire to maximize profits or minimize losses?

A. 5
B. 6
C. 7
D. 8

Form 1 of the Adding/Multiplying Quiz (General Skill Task)

8. The owner of a company must choose between four money schemes. For each scheme, the owner of the firm gets the specified amount. Which scheme has the highest **average** payout? *Hint: to answer this question you'll need to calculate the average for each Scheme -- i.e. to find the average of Scheme 1 you would want to calculate (8+12+4+24)/4.*

Scheme 1	Scheme 2	Scheme 3	Scheme 4
8	4	12	16
12	40	20	16
4	4	4	6
24	12	20	18

   A. Scheme 1
   B. Scheme 2
   C. Scheme 3
   D. Scheme 4

9. A real estate developer has to choose between four apartment buildings. Each building has four types of apartments which charge specific levels of rent. Which building generates the highest revenue (total rent)?

Rent	Building 1	Building 2	Building 3	Building 4
1000	2	4	8	4
1500	4	2	4	6
2000	8	6	2	6
2500	2	4	4	2

   A. Building 1
   B. Building 2
   C. Building 3
   D. Building 4

10. A factory owner must upgrade his factory line. The factory line has five parts. He has bids from four machine shops. Each bid includes all five parts. If the factory owner wants to pay the lowest cost possible for the upgrade, which bid should he choose?

Part	Bid 1	Bid 2	Bid 3	Bid 4
Part 1	10,000	8,000	12,500	9,000
Part 2	12,000	14,000	8,000	11,000
Part 3	1,500	4,500	3,500	4,000
Part 4	2,000	3,000	3,000	4,000
Part 5	16,000	12,000	15,000	13,000

   A. Bid 1
   B. Bid 2
   C. Bid 3
   D. Bid 4

11. Your financial planner presents you with four possible stock portfolios. Each portfolio includes five different stocks, each with a different return. Each portfolio includes 1 of each stock (so there are 5 stocks in total in each portfolio). If you want the highest average return, which portfolio should you choose?

Stock	Portfolio A	Portfolio B	Portfolio C	Portfolio D
Stock 1	20%	25%	15%	10%
Stock 2	10%	10%	10%	12%
Stock 3	25%	20%	15%	20%
Stock 4	6%	9%	6%	12%
Stock 5	4%	6%	14%	6%

A. Portfolio A
B. Portfolio B
C. Portfolio C
D. Portfolio D

12. You have to play one of the four lotteries listed in the table below. If you want to choose the one with the highest average pay out, which lottery should you choose? *Hint: to answer this question you'll need to calculate the average pay-out for each Lottery – i.e. to find the average pay-out for Lottery A you would do the following calculation, (8+16+12+32)/4.*

Lottery A	Lottery B	Lottery C	Lottery D
8	16	24	8
16	20	16	24
12	8	14	14
32	28	22	26

A. Lottery A
B. Lottery B
C. Lottery C
D. Lottery D

13. The owner of a company must choose between four money schemes. For each scheme, the owner of the firm gets the specified amount. Which scheme has the highest **average** payout? *Hint: to answer this question you'll need to calculate the average for each Scheme -- i.e. to find the average of Scheme 1 you would want to calculate (24+24+8+32)/4.*

Scheme 1	Scheme 2	Scheme 3	Scheme 4
24	28	20	20
24	20	18	24
8	6	11	8
32	30	35	32

A. Scheme 1
B. Scheme 2
C. Scheme 3
D. Scheme 4

14. A real estate developer has to choose between four apartment buildings. Each building has four types of apartments which charge specific levels of rent. Which building generates the highest revenue (total rent)?

Rent	Building 1	Building 2	Building 3	Building 4
3000	8	3	5	2
2000	4	6	5	4
2500	6	4	5	8
1500	10	12	11	10

A. Building 1
B. Building 2
C. Building 3
D. Building 4

15. A factory owner must upgrade his factory line. The factory line has five parts. He has bids from four machine shops. Each bid includes all five parts. If the factory owner wants to pay the lowest cost possible for the upgrade, which bid should he choose?

Part	Bid 1	Bid 2	Bid 3	Bid 4
Part 1	30,000	32,000	28,500	29,000
Part 2	10,000	12,000	11,000	9,000
Part 3	22,000	20,000	21,000	20,000
Part 4	12,000	7,000	14,000	9,000
Part 5	4,000	6,000	5,000	11,000

A. Bid 1
B. Bid 2
C. Bid 3
D. Bid 4

16. Your financial planner presents you with four possible stock portfolios. Each portfolio includes five different stocks, each with a different return. Each portfolio includes 1 of each stock (so there are 5 stocks in total in each portfolio). If you want the highest average return, which portfolio should you choose?

Stock	Portfolio A	Portfolio B	Portfolio C	Portfolio D
Stock 1	3%	2%	6%	7%
Stock 2	18%	14%	15%	12%
Stock 3	22%	25%	20%	20%
Stock 4	8%	8%	10%	10%
Stock 5	4%	6%	9%	6%

A. Portfolio A
B. Portfolio B
C. Portfolio C
D. Portfolio D

17. You have to play one of the four lotteries listed in the table below. If you want to choose the one with the highest average pay out, which lottery should you choose? *Hint: to answer this question you'll need to calculate the average pay-out for each Lottery – i.e. to find the average pay-out for Lottery A you would do the following calculation, (150+350+650+350)/4.*

Lottery A	Lottery B	Lottery C	Lottery D
150	100	100	300
350	400	500	400
650	600	550	580
350	300	250	200

A. Lottery A
B. Lottery B
C. Lottery C
D. Lottery D

Appendix B

Table B1: Balance Test: Are Summer Students Different?

	E(NA) (1)	E(A) (2)
A student	0.021	0.003
	(0.032)	(0.045)
Female	0.041	0.022
	(0.056)	(0.050)
Hispanic	-0.043	-0.005
	(0.044)	(0.032)
Black	-0.001	0.001
	(0.012)	(0.014)
Asian	0.035	0.009
	(0.056)	(0.050)
Other	-0.030	0.013
	(0.027)	(0.019)
Accounting	-0.060	-0.43
	(0.054)	(0.047)
Other Major	0.029	0.033
	(0.043)	(0.033)
Sophomore	-0.070	-0.014
	(0.047)	(0.041)
Junior	0.066	0.023
	(0.394)	(0.046)
Senior	0.105**	0.150**
	(0.035)	(0.032)
Age 20-21	0.111**	0.116**
	(0.053)	(0.050)
Age 22 plus	-0.026	-0.032
	(0.043)	(0.041)
Sample Size	514	722

Notes: Each coefficient is from a separate regression where a given covariate is regressed on an indicator equal to one if the academic term is summer. Session fixed effects are included in each regression. E(A) denotes students who expect to earn an A (high expectation) and E(NA) denotes students who do not expect to earn an A (low expectation). A student includes students who earned an A- or above in their intermediate microeconomics course and is defined in Section 5.1. Standard errors are clustered at the session level and are reported in parentheses. ** (*) indicates statistically significance at the 5 (10) percent level. The sample size for A student grade E(NA) and A student grade E(A) is smaller. It is 504 and 709, respectively.

Table B2: Summary Statistics - Test Scores (TUCE) and Expected Grades

	Score	Standardized Score	E(A)	Sample Size
	(1)	(2)	(3)	(4)
Panel A: Treatment Group				
Piece rate	5.50	-0.02	0.57	422
	(2.04)	(1.00)	(0.50)	
Group of Two	5.14	0.02	0.60	642
	(2.06)	(1.00)	(0.49)	
Group of Six WTA	5.10	0.06	0.64	309
	(1.91)	(0.93)	(0.48)	
Group of Six THP	4.76	-0.06	0.64	311
	(1.90)	(0.93)	(0.48)	
Group of Ten WTA	4.81	-0.02	0.65	321
	(1.99)	(1.02)	(0.48)	
Group of Ten THP	5.17	0.03	0.64	299
	(2.06)	(1.05)	(0.48)	
Panel B: Gender				
Male	5.36	0.12	0.64	1,381
	(2.02)	(1.00)	(0.48)	
Female	4.73	-0.17	0.58	923
	(1.96)	(0.95)	(0.49)	
Panel C: Class Standing				
Lower Division	4.84	0.000	0.61	1,678
	(1.97)	(0.99)	(0.49)	
Upper Division	5.82	0.03	0.64	626
	(1.99)	(0.99)	(0.48)	

Notes: Mean scores and expected grades are reported by subgroups. Standard deviations are in parentheses. E(A) denotes students who expect to earn an A (high expectation) and E(NA) denotes students who do not expect to earn an A (low expectation). WTA are treatments where the top scorer is paid, THP are treatments where the top half of the group is paid.

Table B3: Pre-Treatment Characteristics - Sample Balance

		Differences Between Treatments				
	Group of 2 Mean (1)	Piece Rate (2)	Group of 6 WTA (3)	Group of 6 THP (4)	Group of 10 WTA (5)	Group of 10 THP (6)
Panel A: E(NA)						
Female	0.41	-0.04	0.14	0.11	0.20	0.04
	(0.49)	[0.49]	[0.06]	[0.17]	[0.03]	[0.65]
Hispanic	0.18	0.06	-0.09	0.04	0.08	-0.10
	(0.39)	[0.23]	[0.12]	[0.47]	[0.27]	[0.15]
Black	0.02	-0.01	0.04	0.00	0.00	0.04
	(0.12)	[0.60]	[0.06]	[0.96]	[0.94]	[0.12]
Asian	0.37	-0.09	0.02	-0.09	-0.05	0.03
	(0.48)	[0.18]	[0.78]	[0.25]	[0.57]	[0.71]
Other	0.07	-0.01	-0.03	-0.04	-0.06	-0.03
	(0.26)	[0.70]	[0.45]	[0.32]	[0.20]	[0.42]
Accounting Major	0.42	0.09	0.03	-0.05	-0.05	-0.15
	(0.49)	[0.14]	[0.70]	[0.50]	[0.55]	[0.06]
Other Major	0.30	-0.09	0.01	0.09	0.10	0.07
	(0.46)	[0.07]	[0.83]	[0.12]	[0.14]	[0.31]
Sophomore	0.30	0.11	-0.11	0.11	0.15	-0.04
	(0.46)	[0.03]	[0.06]	[0.10]	[0.06]	[0.57]
Junior	0.36	-0.07	0.11	0.06	0.01	0.06
	(0.48)	[0.22]	[0.06]	[0.38]	[0.88]	[0.39]
Senior	0.19	0.00	-0.09	-0.04	-0.05	-0.10
	(0.39)	[0.95]	[0.02]	[0.29]	[0.21]	[0.04]
Age 20-21	0.38	0.05	0.05	0.08	0.05	-0.04
	(0.49)	[0.45]	[0.47]	[0.24]	[0.49]	[0.58]
Age 22+	0.17	-0.04	-0.05	-0.05	-0.07	0.00
	(0.37)	[0.40]	[0.35]	[0.37]	[0.23]	[0.96]
Sample Size	256	439	367	369	367	364
Panel B: E(A)						
Female	0.39	-0.09	0.02	-0.08	0.03	0.02
	(0.49)	[0.10]	[0.69]	[0.15]	[0.63]	[0.77]
Hispanic	0.14	-0.03	0.06	-0.06	-0.04	0.03
	(0.35)	[0.42]	[0.14]	[0.13]	[0.32]	[0.52]
Black	0.02	0.00	0.00	0.01	0.00	-0.02
	(0.13)	[0.88]	[0.83]	[0.54]	[0.73]	[0.24]
Asian	0.45	0.00	0.00	0.04	0.01	-0.01
	(0.50)	[0.95]	[0.95]	[0.47]	[0.93]	[0.90]
Other	0.05	-0.03	0.01	0.01	0.02	0.02
	(0.21)	[0.20]	[0.64]	[0.66]	[0.51]	[0.63]
Accounting Major	0.42	-0.01	-0.07	0.04	0.00	0.08
	(0.49)	[0.84]	[0.21]	[0.45]	[0.99]	[0.17]
Other Major	0.29	0.08	0.06	-0.04	-0.06	-0.04
	(0.46)	[0.05]	[0.19]	[0.33]	[0.09]	[0.41]
Sophomore	0.27	-0.05	-0.05	-0.04	0.03	0.01
	(0.45)	[0.24]	[0.24]	[0.38]	[0.58]	[0.90]
Junior	0.39	0.04	0.03	0.02	-0.10	0.02
	(0.49)	[0.43]	[0.52]	[0.73]	[0.04]	[0.78]
Senior	0.19	0.03	-0.04	-0.02	0.04	-0.04
	(0.39)	[0.34]	[0.18]	[0.60]	[0.29]	[0.36]
Age 20-21	.39	0.06	0.11	-0.02	-0.11	0.06
	(0.49)	[0.29]	[0.04]	[0.68]	[0.04]	[0.36]
Age 22+	0.22	0.01	-0.11	-0.01	0.02	-0.10
	(0.41)	[0.84]	[0.01]	[0.72]	[0.69]	[0.06]
Sample Size	386	625	584	584	599	577

Notes: Column 1 reports means for the group of two treatment. Corresponding standard deviations are in parentheses. Differences in columns 2-6 are from OLS regressions that include session fixed effects. Sample sizes listed in columns 2-6 include the group of two and the group listed at the top of each column. P-values for differences are in square brackets. E(A) denotes students who expect to earn an A (high expectation) and E(NA) denotes students who do not expect to earn an A (low expectation). WTA are treatments where the top scorer is paid, THP are treatments where the top half of the group is paid.

Table B4: Score Differences by Incentive Structure Across Grade Expectation

	Difference Estimates		Effects Statistically Different?
	E(NA) (1)	E(A) (2)	(1)-(2) (3)
Group of Two - Piece Rate Difference	-0.164*	0.061	-0.225**
	(0.088)	(0.088)	(0.107)
Group of Six WTA - Piece Rate Difference	-0.322**	-0.032	-0.290
	(0.133)	(0.146)	(0.187)
	[0.13]	[0.46]	
Group of Six THP - Piece Rate Difference	-0.270**	-0.056	-0.212
	(0.126)	(0.124)	(0.140)
	[0.29]	[0.22]	
Group of Ten WTA - Piece Rate Difference	-0.202	0.060	-0.262*
	(0.155)	(0.122)	(0.153)
	[0.78]	[0.99]	
Group of Ten THP - Piece Rate Difference	-0.384**	-0.016	-0.369**
	(0.163)	(0.150)	(0.137)
	[0.11]	[0.55]	
Equivalence of Group Coefficients [P-Values]	0.36	0.54	

Notes: Sample size is 2,304. All coefficients are from a single regression (equation 1) that includes indicators for race, gender, major, and year in school, as well as session fixed effects. E(A) denotes students who expect to earn an A (high expectation) and E(NA) denotes students who do not expect to earn an A (low expectation). WTA are treatments where the top scorer is paid and THP are treatments where the top half of the group is paid. Standard errors are clustered at the session level and are reported in parentheses. Square brackets report the p-value testing for the equivalence of the corresponding coefficient and the coefficient for the group of two. ** (*) indicates statistical significance at the 5 (10) percent level.

Table B5: Score Differences by Incentive Structure Across Grade Expectations by Gender

	E(NA)		E(A)	
	Male	Female	Male	Female
	(1)	(2)	(3)	(4)
Group of Two - Piece Rate Difference	-0.227*	-0.071	0.110	-0.023
	(0.114)	(0.150)	(0.106)	(0.141)
Large Group WTA - Piece Rate Difference	-0.432***	-0.194	-0.045	-0.023
	(0.149)	(0.167)	(0.139)	(0.148)
	[0.07]	[0.38]	[0.13]	[1.00]
Large Group THP - Piece Rate Difference	-0.267*	-0.266	0.093	-0.119
	(0.144)	(0.166)	(0.140)	(0.156)
	[0.75]	[0.15]	[0.85]	[0.52]
Equivalence of Group Coefficients [P-Values]	0.20	0.35	0.26	0.70

Notes: Sample size is 2,304. All coefficients are from a single regression (equation 1) that includes indicators for race, gender, major, and year in school, as well as session fixed effects. E(A) denotes students who expect to earn an A (high expectation) and E(NA) denotes students who do not expect to earn an A (low expectation). WTA are treatments where the top scorer is paid and THP are treatments where the top half of the group is paid. The 6 and 10 person WTA (6 and 10 person THP) tournament treatments are collapsed into a single group category denoted as Large Group WTA (Large Group THP). Standard errors are clustered at the session level and are reported in parentheses. ** (*) indicates statistically significance at the 5 (10) percent level.