Advance Behavioral Economics

Zainab Azhar

Saad Siddique

Zain Ayaz

Reema Hajat

Zunaira Khan

Research Question:

What are the changes in behavior dynamics, in terms of willingness to help peers, under different grading system: absolute and relative?

Contents

Introduction and Research Question

There is a long-standing debate on whether absolute or relative grading systems are better. Both systems have different perks, flaws, and effects on the behavior of students. Each of these two grading systems provides different incentives to students, which in turn influences them to alter particular aspects of their behavior, such as effort, productivity, cooperation, and so on, to score well under the respective system. Many studies have been conducted to capture differences in productivity of students when being graded under these two opposing systems. Generally, relative grading systems incentivize greater effort in studying and attempting exams than absolute grading systems. Conversely, students are more cooperative with and helpful towards their peers under absolute grading systems. Our focus is to examine the way helpfulness towards peers changes under relative grading systems, where we expect students to be more reluctant to help others under fierce competition. We hypothesize that an average student will be most reluctant to help those with much higher ranks, and more likely to help those with low ranks compared to their own. This effect, caused by the change in incentives under fiercer competition, has been attributed to risk aversion by many studies. We argue that this change in behavior is not completely explained by risk aversion, but other-regarding preferences of students, in the form of inequity aversion, also play a significant role. Hence, we will be using the theoretical basis of the Fehr-Schmidt model to examine the changes in other-regarding preferences caused by a change in the grading system. We predict that beta, the parameter of advantageous inequity (altruism), will be lower, on average, and alpha, the parameter of disadvantageous inequity (envy), will be higher for the class evaluated by relative grading. Thus, our research question is as follows:

What are the changes in behavior dynamics, in terms of willingness to help peers, under different grading systems: absolute and relative?

Literature Review:

The Journal Article titled "The effect of using relative and absolute criteria to decide students' passing or failing a course" by Ayfer Sayin found that the "relative criterion" (relative grading) is more efficient than "absolute criterion" (absolute grading) in terms of "letter grading" as well as student's decision-making process of passing a course. The criterion in this scenario is a pre-calculated "absolute" value which does not depend on the group and applies to everyone. Relative evaluation is a type of evaluation that is based on a criterion derived from measurement findings. The grades associated with letters for graduate students who were passing their courses at Gazi University used the relative grading system. However, absolute criteria are meant to determine whether the course has been passed or failed of the same course, even if the course is most of the time taught by the same instructor. "The research group in this study consisted of 141 students who received formation education in two different classes at Gazi University during the academic year 2015-2016" (Sayin, 3). Out of these students, 73 were studying Turkish philosophy and 68 were studying physical education. The research data were gathered using the researcher's created measurement and assessment midterm and final tests. "Each achievement test had 25 multiple-choice questions with five different solutions" (Sayin, 5). It was seen that the students who score below 42 in relative criterion were supposed to fail the course. While in absolute criterion below 60 were supposed to fail the course. "It was observed that the number of students with high grades of AA, BA, and BB for the relative evaluation, is higher than in the absolute evaluation" (Sayin, 6). Even though this paper has compared two different courses of the same students, but it is useful for our paper because it has proved the point that grades in

relative criteria are higher than in absolute grading. Our model is also reinforcing this point that due to competition and effort by student in relative grading system, students tend to receive higher grade.

An experiment by Taylor Burleigh and Danial Meegan shows how students behave in an absolute grading setting and normative (relative) setting and peer-status proximity whether the students' status is proximate, distant or unknown. They used two prospects: risky prospect and risk aversion to determine the peer-status and their behavior towards assistance respectively. The research and experiment show how the behavior of mastery-oriented individuals towards helping their peers differ from performance-oriented individuals in normative settings. Students are less likely to help others in normative and if the person is performance-oriented, she/he less likely to help others and also can provide low quality assistance, he will only focus on out performing their peers. So, it won't hurt their own grades. On the other hand, mastery-oriented students prefer to master one task at a time and focus on learning and improvement in an absolute grade setting. There's a zero sum issue in normative grading which means the gain of one student is the loss of another, so this is also one of the reasons why students don't help their peers. Normative grading and peer-status form risky prospects. It doesn't mean students are completely competitive and don't help at all in a relative setting, so it's a "mixed-motive interdependence" in which there are zero-sum and non-zero-sum situations for students. Individuals' status do play an important part. For unknowns in normative settings tend to be risk averse and helping others could backfire.

Burleigh and Meegan used the sample size of 234 students (97 females and 137 males) to determine the impact of grading setup on peer-status whether they'll help or not and competitiveness. In the methodology, they used normative grading vignette and absolute grading

vignette to check the policy impact and for status impact, they use proximate status vignette and distant status vignette.

In conclusion, students with similar peer-status are less-cooperative in a normative setting and students in an absolute environment are more likely to help. In the short term, normative could improve skills through competitiveness while it also induces misfit culture in students and not encouraging teamwork. While, in absolute grading setup, students are more likely to help which can improve learning, bond with peers, and promotes teamwork.

Another paper titled “A classroom experiment on effort allocation under relative grading” by Andy Brownback is about how size of a class impacts the effort levels of the students. A classroom based experiment about the relative grading system has been conducted in an economics class at the “University of California, San Diego (UCSD)”. “Two quizzes were assigned, one to a cohort of 10 students, called 10-student quiz, and the other to a cohort of 100 students, called 100-student quiz” (Brownback, 114). It was realized that the efforts in the 100-students quiz was 3% more than in the 10-students quiz and this difference was statistically significant. Effort is measured by how much time is spent preparing for the quiz. Also, GPA data was used to determine the impact of size of the cohort on students having dissimilar potentials. “The model predicts that larger cohorts reduce low-ability students' incentives for effort while increasing high-ability students' incentives for effort” (Brownback, 115). Furthermore, the framework speculates that as larger cohorts' uncertainty decreases, aggregate effort would rise. It is useful for our model to see how relative grading affects the level of effort in different abilities of students. The literature review allows us to define absolute and relative grading. Based on the findings of these paper, we can support our predictions with valid findings regarding the advantages and disadvantages of each system. The finding that competitiveness is observed to be

generally higher under the relative grading system is what will serve as the basis for our attempt to explain this change in behavior by incorporating the role of other-regarding preferences into our model. We want to test that risk aversion, which has been found by many studies, is not the sole factor behind the increase in competitiveness. We predict that other-regarding preferences (inequity aversion) in the form of increased envy, and lower altruism also play a role. While relative grading percentages can vary, an absolute scale stays fixed. Thus ranks indicate how well you are doing better than a percentage alone under relative grading. A typical absolute grading scale is as follows:

Grade	SUS	Percentile range
A+	84.1 - 100	96 - 100
A	80.8 - 84.0	90 - 95
A-	78.9 - 80.7	85 - 89
B+	77.2 - 78.8	80 - 84
B	74.1 - 77.1	70 - 79
B-	72.6 - 74.0	65 - 69
C+	71.1 - 72.5	60 - 64
C	65.0 - 71.0	41 - 59
C-	62.7 - 64.9	35 - 40
D	51.7 - 62.6	15 - 34
F	0 - 51.6	0 - 14

Fig 1 An absolute grading scale does not depend on competition between students

Theoretical Motivation

Traditional neoclassical economics assumes that an individual is rational and self-regarding. However, empirical evidence shows that people have other-regarding preferences, such as envy and altruism, and the weight of these factors affects their utility. For example, the Fehr Schmidt model describes self-centered inequity aversion, which incorporates weights for the degree of altruism and envy a person feels for others who are not earning at the same level as

them. This model also encompasses the case of pure self-regarding preferences when both parameters for envy and altruism are zero. In that case, alpha and beta both would be equal to zero, and only the level of one's own wealth would determine the utility derived from that level of income. However, when people are envious of those earning higher than themselves, they experience a level of disutility at their current level of wealth. If one feels empathetic towards those earning less than him, they will experience disutility. So, the overall utility will fall if the value of either of these parameters is positive.

The Fehr Schmidt model of inequity aversion is as follows:

$$U_i(y_i, y_{-i}; \alpha_i, \beta_i) = y_i - \frac{\alpha_i}{n-1}\sum_{j \neq i} \max\{y_j - y_i, 0\} - \frac{\beta_i}{n-1}\sum_{j \neq i} \max\{y_i - y_j, 0\}$$

According to this theory, the parameter of envy, alpha, is greater than or equal to zero. Beta, the parameter of altruism, is greater than or equal to zero, and less than one. As verified by empirical tests, beta is assumed to be less than alpha, meaning the disutility derived from people doing better than oneself is worse than that of people doing worse than oneself. In short, the disutility of envy is stronger than the disutility of feeling sympathetic and guilty. Other regarding preferences exist if at least one of these parameters is positive. This model (henceforth referred to as the FS model) allows for non-linear utility functions. The size of the reference group, N, is not restricted or specified. To accommodate the size of the group of players, inequalities are normalized (Dhami).

Instead of income, we apply this model to grades, in terms of ranks, which are obtained from different systems. Evidence, as explained in the literature review, suggests that students

become more competitive under the relative grading system, but are more cooperative with peers under the absolute grading system. To model this behavior, we have introduced a variation of the FS model, which is explained under the *Empirical Methodology and Data Section* below.

Empirical Methodology and Data

Data

One course at LUMS with large section sizes (e.g., Principles of Macroeconomics, Calculus 1. etc.) will be chosen. In that particular semester, one section will be graded on an absolute scale, while another will be graded on a relative scale. Of course, there will be differences between individual students within and between the groups in terms of ability, wealth, privilege, IQ, effort, and other factors. However, we are choosing two sections of the same course with the same instructor in order to ensure that the two groups, on average, are the same across all factors. These two groups must be systematically the same to draw a valid comparison. Thus, any differences found between the two groups will be because of the difference in grading systems, and not because of any other external factors. Our aim is to explain this difference in terms of other-regarding preferences, though we do recognize that risk aversion is another factor which gives rise to this difference. We choose the largest possible sample size possible by using a course with large sections because, as explained in the literature review, the difference in competitiveness is more prominent and easier to capture, the larger the sample size. Similarly, we are trying to maximize our chances of picking up the existing differences in the degree of helpfulness under the two different grading systems. The control group is the one with absolute grading, and the treatment group is the one with relative grading.

First, the two sections will be notified of their respective grading systems in their respective first sessions, as instructors usually do when going over the course outline. Midway through the semester, the two classes will take their first graded assessment: a midterm exam. Both sections, separately, will be told that their rankings in the class will be privately sent to them after the midterms are graded, and they need to answer an anonymous, ungraded survey after that. After the midterms are marked, both sections will be called to separate classrooms, with their respective TA's invigilating them. They will be told to sit with a one-seat space between them to fill an anonymous survey, before the regular session starts. Students are not allowed to talk until this whole process is over. The TA will announce that he/she has received everyone's ranking in terms of how well they scored from the Professor. The TA will call each student one by one and show them a sheet with their roll numbers on it and their own rank. However, this will be a false announcement because every student's sheet will notify them that they are at the 25th rank (where 51 students are in each section) which represents the mean mark. No student at that moment will know that everyone has been notified as having the same mean rank and will assume this is their own rank. Then, an anonymous form will be passed out to everyone which will have the following questions and information:

> A few students in your class have missed the crucial lecture which will be included in your final exam. They need notes to be able to solve a significant part of the final exam.
>
> 1. Given your rank, would you be willing to lend your notes and/or provide additional help to someone who attained the 3rd rank in the midterm?
>
> 2. Given your rank, would you be willing to lend your notes and/or provide additional help to someone who attained the 15th rank in the midterm?

3. Given your rank, would you be willing to lend your notes and/or provide additional help to someone who attained the 23rd rank in the midterm?

4. Given your rank, would you be willing to lend your notes and/or provide additional help to someone who attained the 27thrank in the midterm?

5. Given your rank, would you be willing to lend your notes and/or provide additional help to someone who attained the 35th rank in the midterm?

6. Given your rank, would you be willing to lend your notes and/or provide additional help to someone who attained the 47th rank in the midterm?

As a shortcut when making decisions and forming perceptions, people tend to relate the information they first perceived to the following situations/information, causing their answers to these questions changing because of the ordering of these questions and the information they provide. This is known as the anchoring heuristic, and to avoid this creating bias in the results, these questions will be randomly shuffled in each sheet. In the questions, we made the difference between the set rank of 25th out 51 and the given rank of student who needs help gradually increase. In some questions, the distance of the comparison rank is small, then medium, followed by large. The answer scale will be of Likert form, which is widely used by researchers to capture opinions and attitudes of participants. The Likert form works best when there are more than 5 and less than 11 options (Leung). It will capture the willingness to help.

Options	Not willing at all	Not willing	More or less not willing	Indifferent	More or less Willing	Willing	Completely willing

Number Scale	1	2	3	4	5	6	7

We will quantify the responses from the questionnaire into a number scale. Then, we will calculate an average of each question's responses separately, which will give us 6 distributions for each section. This will be done for both sections. Next, we will plot to see, on average, at what point students shift from the range of willingness to help to unwillingness to help.

Empirical Methodology

Explaining the BehWllingness to Help of Students With Using an Altered FS Model:

Hypothesis testing, as explained in the falsification test section below, will be conducted to determine if students behave differently in absolute and relative grading in terms of willingness to help. The difference in their preferences to help or not help can be due to several reasons, but we will try to explain this using Fehr-Schmidt model of other-regarding preferences.

Our response variable is the score from the Likert scale which respondents (students) filled out for 6 different questions about willingness to help other students of differing ranks. Students will help if they have other regarding preferences. We will explain this behavior using our altered verion of the FS model, as follows:

$$U_i(y_i,y_{-i},a_i,b_i) = a_o - a_i/(n-1)(\text{sum}(\max\{y_j - y_i,0\})) + b_i/(n-1)(\text{sum}(\max\{y_i - y_j,0\}))$$

where,

U_i is the average response of a particular student to 6 questions (has a range from 1 to 7).

a_o is the intercept term

a_i is the parameter of envy of a particular student

b_i is the parameter of altruism of a particular student

y_i is the rank of a particular student

y_j is the rank of the students other than i

In comparison with the original FS model, we have used positive sign with advantageous inequity. The reasoning behind this can be explained with an example. Suppose the original FS model equation is being used to derive the overall utility of an individual "i" from his income, y_i. If this person has other-regarding preferences, he may feel envious of those with higher earnings than him, such as individual j earning income y_j. He may also feel bad for those earning less than him, where the y_j corresponding to beta is the earnings of one such individual, j. Since we are using ranking instead of income, the lower the numerical value of the rank, the higher is the rank. Conversely, in terms of income, the opposite is true. The higher the numerical value of the income, the higher is the income. Thus, we have replaced the minus sign with a plus sign, since the disutility from having other regarding preferences should decrease the overall utility.

Here a and b are positive numbers. Under a relative ranking system, we predict that incentives under a relatively graded system, compared to an absolute grading system, will change other-regarding preferences of students such that $a_i > b_i$. This implies that the disutility from disadvantageous inequity will be greater than the disutility that comes from advantageous inequity because of fiercer competition. We do not claim that increasingly competitive behavior is purely explained by changes in other-regarding preferences; other factors, such as can also partly explain this phenomenon. This elaborated on under the section *Limitations*.

After collecting the data from the answers of each section as previously described, we would observe that, despite the variables of rank differences ($\mathbf{y_j - y_i}$) and ($\mathbf{y_i - y_j}$) remaining the same, there will be a change in willingness to help i.e., difference in answers to the questions between the two sections. We will take an average of all responses of those in the section with relative grading and regress it will respect to the two variables as shown in the above equations. This will result in an average a (alpha) and average b (beta) for this section. We will repeat the same for absolute grading class students as well and find an average alpha and beta. Then we will compare alphas and betas with each other.

We Predict that:

1. The alpha for relatively graded section will be greater than or equal to the alpha for section with an absolute grading system, since students in the absolute section will not care about the rank of their fellow students, so the degree of envy observed may be less intense than that of the relative case.
2. The beta for relative case will be greater than the beta for absolute case (Since relative class students suffer more if someone is above in rank)

Falsification Tests:

Our methodology includes hypothesis testing, both within the sections and across the sections. Since we will have 6 different distributions stemming from the 6 questions, there will be 120 different combinations. To separate observed variance data into different parts, we will use ANOVA hypothesis testing. The ANOVA test will determine the significance of the results, and hence tell us if the two groups are different from each other in terms of willingness to help,

via the p-value with a significance level of 5%. Also, we will determine whether there are significant differences in the respective distributions of each section, in terms of the sample means, modes, and standard deviations. When comparing across groups, we will test the hypothesis for each of the questions separately, so there will be a total of 6 hypothesis tests. The mean for each of the 6 question responses will be taken and stored in six different variables for each of the two sections. We will then have a total of 12 average variables (6 for each section). We will then test the difference between the mean values of each question in the section with relative grading with that of the respective question in section with absolute grading.

Our null and alternative hypotheses are as follows:

H0 (Null Hypothesis): There is no difference in the distributions

H1 (Alternate Hypothesis): There will be difference in the distributions

If H0 is rejected, it means at least one question has a different distribution from other questions. If this difference is significant, then this indicates that there was a change in behavior of students. In terms of willingness to help other peers, under the relative grading system as compared to the absolute grading system. Based on the mean value, we can assign each group (section) a point from the Likert scale. If the students in the section with absolute grading have a higher mean value for the questions, provided that the difference between their mean value and the relatively graded section's mean value for the question is significant, we can say students in the sections with absolute grading are more willing to help than the ones in graded course. This could be because those in the absolutely graded section do not care about rank as much as those under relative grading since the competition factor is eliminated, and so they would answer

based on their true envious/altruistic tendencies. In a relatively graded section, there is more competition, which may fuel more envy and less altruism.

Conclusion

Policy Implications

We do not claim that increasingly competitive behavior is purely explained by changes in other-regarding preferences; other factors, such as risk aversion, also partly explain this phenomenon. Many studies have focused on risk-aversion and other factors as explanations, we want to research into the role of other-regarding preferences to create a broader picture for further research which can incorporate and tease apart the effects of several different explanations. Our results can be used as a starting point for further studies regarding the main goals of a university. If proof that relative grading systems make people more envious and less altruistic is found, we still do not advocate that one particular system should be used over the other because each system has its own advantages and disadvantages. If the goal of the university is to promote cooperation and helpfulness among peers, then absolute grading may be the ideal choice. However, if the goal is to foster competitiveness and increased efficiency among students, then relative grading may be the ideal choice. Also, the degree of difference between the alpha and betas of the absolute system and the relative system may indicate how easy/difficult it is to alter students' other regarding behavior. If the goal is to develop students into helping and cooperating individuals, this information can be used to develop alternative policies for the classroom that may bring about a more persistent, and ingrained change in

students' helping behavior despite opposing incentives they may face in the real world. This is also an interesting point for further research.

Causality and Limitations

1. In this paper, we had all of the students believe they are of the same rank. However, it is possible that the unwillingness of the student at the 1st rank to help a student at the 3rd rank may be more than that of the students of the 12th rank and 14th rank. In both scenarios, the student that needs help is two ranks below the one who can help, but the level of competitiveness at different ranks may not be the same. Moreover, students with a large difference between their ranks may be quite different in characteristics, privileges, and ability. Our study can prove helpful to the people who are willing to construct a more complex model that incorporates these factors.
2. The differences found can also be due to risk aversion in their behavior. Further complex studies can be done to tease apart the influence of other-regarding preferences and risk aversion in this situation. Risk aversion would be that the students do not want to risk dropping in ranks, and hence their grades, in a relatively graded section. Here we accounted for others regarding preferences. Hence, we do not claim that any significant differences found in this study are solely attributed to other-regarding preferences. There are other influences on preferences, such as risk aversion, involved.

Citations

Brownback, Andy. “Classroom experiment on effort allocation under relative grading”. Economics of Education Review 62 (2018) 113–128

Burleigh,T.J., Meegan,D.V. “Risky prospects and risk aversion tendencies: does competition in the classroom depend on grading practices and knowledge of peer status?”. Soc Psychol Edu 21, 323-335(2018).

Leung, Shing-On. “A Comparison of Psychometric Properties and Normality in 4-, 5-, 6-, and 11-Point Likert Scales.” *Taylor & Francis Online*,

https://www.tandfonline.com/doi/full/10.1080/01488376.2011.580697?casa_token=8YAqC4hjHh8AAAAA%3Am9qbTqULLWjC9Hh7TGzFCRLNjingfqmLcP9WZch2RjVRGtbavlCsHIqqfdmh_zEI-niJ3vJ1gOtmONz2.

Sanjit Dhami (2016) Foundations of Behavioral Economic Analysis. Oxford: Oxford University Press

Sayin, Ayfer. “The effect of using relative and absolute criteria to decide students”. Journal of Education and Training Studies, v4 n9 p1-9 Sep 2016