

Discussion Paper Series

Graduate School of Economics and School of Economics

Meisei University

Discussion Paper Series, No.40

July, 2018

**Information Feedback in Relative Grading: Evidence from a Field
Experiment**

Shinya Kajitani

(Kyoto Sangyo University)

Keiichi Morimoto

(Meisei University)

Shiba Suzuki

(Seikei University)

Hodokubo 2-1-1, Hino, Tokyo 191-8506

School of Economics, Meisei University

Phone: +81-(0)42-591-9479 Fax: +81-(0)42-599-3024

URL: <https://keizai.meisei-u.ac.jp/econ/>

E-mail: keizai@econ.meisei-u.ac.jp

Information Feedback in Relative Grading: Evidence from a Field Experiment

Shinya Kajitani* Keiichi Morimoto[†] Shiba Suzuki[‡]

July 3, 2018

Abstract

The impact of relative performance information feedback could vary according to each student's previous examination performance. Binary grade environments enable us to identify the heterogeneous impacts of this feedback. Conducting a randomized control trial employing a compulsory course in economics at a Japanese university, we show the heterogeneous impacts of relative performance information feedback attributable to the students' earlier examination scores under a binary grade environment. Our experimental results prove that previous performance information feedback improves the performance of students with only intermediate scores but worsens the performance of high-scoring students in their next examination.

Keywords: Education; experiments; relative performance information feedback; relative grading.

JEL Classification: D81, I21

*Corresponding author. Associate Professor, Faculty of Economics, Kyoto Sangyo University, Motoyama, Kamigamo, Kita-ku, Kyoto-city 603-8555, Japan. Phone number: +81-75-705-1739. E-mail address: kajitani@cc.kyoto-su.ac.jp

[†]Associate Professor, Department of Economics, Meisei University, 2-1-1, Hodokubo, Hino-city, Tokyo 191-8506, Japan. Phone number: +81-42-591-6964. E-mail address: keiichi.morimoto@meisei-u.ac.jp

[‡]Associate Professor, Faculty of Economics, Seikei University, 3-3-1, Kichijoji-kitamachi, Musashino-city, Tokyo 180-8633, Japan. Phone number: +81-422-37-3571. E-mail address: shiba.suzuki@econ.seikei.ac.jp

I Introduction

How does relative performance information feedback affect the incentive of students to study for multiple examinations in a relative grading environment? Many consider information feedback associated with a reward scheme as an efficient way of eliciting the incentives of students to study. Further, “relative grading” or “grading on a curve” is widely used in grading students, although its use remains controversial.¹ In relative grading, student grades depend on their positions in the class score distribution. To understand student incentives in a relative grading scheme, Becker and Rosen (1992) extend the rank-order tournament model of Lazear and Rosen (1981) to emphasize the importance of a student’s position in the distribution of academic attainment and show that competition between students indeed stimulates learning effort. This suggests that relative performance information feedback affects student decision-making in providing effort. In addition, in actual schooling environments, students are graded using multiple examinations. It is then worth considering the relationship between information on a student’s relative position in the distribution of earlier examination scores and the student’s incentive to provide study effort for the following examination.²

The impact of relative performance information feedback may vary according to each student’s previous examination performance. In a relative grading scheme, a student needs to receive a higher score than her opponent to obtain a better grade. That is, an opponent’s score serves as a threshold she must exceed. In this grading environment, the relative performance information feedback is then a signal of the effort she should provide. For example, when relative performance information feedback tells the student her current score is relatively low, she understands that she has to provide a higher level of effort to get above the threshold. Conversely, she may give up, saving the cost of effort. In a multiple examinations environment, for students with relatively low marks in the previous examination, relative performance information serves as a signal that they will need to work harder for the next examination to obtain a better grade. In contrast, for students with sufficiently high marks in the previous examination, relative

¹Volokh (2015) argues that teachers appreciate relative grading as a means to control grade inflation and to ensure students have an opportunity to receive a higher grade, even if the examinations make it difficult to obtain high marks. Conversely, Grant (2016) points out that relative grading can be problematic in that it prevents students from collaborating owing to the overly competitive environment.

²Aoyagi (2010) and Ederer (2010) theoretically analyze information feedback in a dynamic tournament context.

performance information may be a signal that they can slack off. Therefore, the impact of relative performance information feedback varies according to each student's previous examination performance. That is, there are heterogeneous impacts associated with previous examination performance.

Binary grade environments, such that only a pass or fail matters, enable us to identify the heterogeneous impacts of the relative performance information feedback on the students' incentive to study according to their previous examination performance. In most academic courses, there are multiple grades, such as A, B, C, D, and F, such that when a student is above the threshold between B and C, she is also below the threshold between A and B. Suppose that the relative performance information feedback has a positive impact on students above the threshold but negative impacts on students below the threshold. That is, the negative aspects of relative performance feedback may cancel out its positive aspects on a given student's incentive to study. In contrast, under binary grade environments, we can accurately identify the effect of relative performance information feedback on the student's incentive to study.³

Conducting a randomized control trial employing a compulsory course in economics at a Japanese university, this paper identifies the heterogeneous impacts of relative performance information feedback due to the students' midterm scores. As explained later, empirical and anecdotal evidence concerning both the university advancement rate and labor markets in Japan suggest that students care only about whether they can pass the course. In our experiment, students receive relative grades using the results of their midterm and final examinations. Given the students are aware of this situation, we can indeed identify the impact of relative performance information feedback in a binary grade environment. We allocated more than 200 students into a control group and a treatment group immediately following the midterm examination. We only provided relative performance information feedback to students in the treatment group and explored the impact of this feedback on student performance in the final examination. We then constructed a simple theoretical model of binary grading to clarify the impact of the rel-

³Binary grade environments are themselves of some practical importance. For instance, some argue that there is grading on a curve in bar examinations in some US states (Stenson, 2017). For example, the Massachusetts Court System states that "A Final Total Scale Score of 270 or greater is required to pass the Massachusetts Bar Examination" (Massachusetts Court System, 2017). In calculating scale scores, there is also sometimes an adjustment of raw scores to account for variations in the level of difficulty of an examination over time.

ative performance information feedback on the student’s incentive where there are many participants. Subsequently, using data from the field experiment, we demonstrated that for students with sufficiently high marks in the midterm examination, students in the treatment group tended to receive lower scores in the final examination than students in the control group. In contrast, for students with intermediate marks in the midterm examination, students in the treatment group tended to earn higher scores in the final examination than students in the control group. These findings are almost consistent with our simple theoretical model.

To our knowledge, this paper is the first attempt to investigate the impact of relative performance information feedback on student incentives to study in an actual educational environment encompassing relative grading. The remainder of the paper is organized as follows. Section II relates the analysis to the existing literature and Section III describes the experimental design. Section IV details the theoretical tournament model and provides the testable hypothesis. Section V presents the empirical framework and reports the estimation results. Section VI concludes.

II Background

Revealing the role of relative performance information feedback in providing agent incentives is often through laboratory experiments. Existing research demonstrates that relative performance information feedback can have both a positive and a negative impact on the agent’s effort provision. For instance, Azmat and Iriberry (2016) and Gill et al. (2018) conduct a laboratory experiment to examine the impact of relative performance information feedback on the subjects’ performance when rewards are absolute, but independent of the other subjects’ performance. In particular, Gill et al. (2018) find that the rank-response function is U-shaped, that is, subjects increase their effort most in response to relative performance information feedback when they are ranked first or last. On the other hand, Eriksson et al. (2009), Freeman and Gelber (2010), and Ludwig and Lünser (2012) conduct laboratory experiments to explore the impact of relative performance information feedback on agent performance when subjects are rewarded relatively. Eriksson et al. (2009) and Freeman and Gelber (2010) conclude that relative performance information feedback lowers the performance of subjects whose interim performance is relatively low. However, those subjects whose midterm performance is relatively high do

not slacken off. In contrast, Ludwig and Lünser (2012) examine the effects of effort information in a two-stage rank-order tournament. They demonstrate that subjects who lead tend to lower their effort, but those who lag increase it relative to the first stage, while the subjects who lead exert a greater effort than those who lag. These results suggest that the impact of relative performance information feedback should vary according to the subjects' initial level of attainment in our relative grading framework.

Information feedback is also a subject of investigation in actual educational decision-making.⁴ Azmat and Iriberry (2010), using data from Spanish high schools, and Tran and Zeckhauser (2012), in a field experiment of Vietnamese university students, demonstrate that relative performance information feedback raises the performance of students when they are rewarded absolutely. Both these studies argue that if students have competitive preferences, which means that they inherently prefer receiving a higher rank than others, relative performance information has a positive impact on their incentive to study, even if the information is intangible in the reward scheme. Bandiera et al. (2015) employ a sample of students rewarded absolutely in a UK graduate school and demonstrate that absolute (but not relative) performance information feedback improves the performance of students whose midterm performance is high, which the authors relate to knowledge of the educational production function. These findings imply that relative performance information feedback may have a substantial impact on student performance, even in our relative grading setup. However, there has been no examination of the impact of relative performance information feedback on student incentives under relative grading in an actual education environment.

Relative rewarding and grading on a curve are themselves worth considering because existing research demonstrates that relative grading in particular has an empirically debatable impact on student performance.⁵ For example, Ashraf et al. (2014) conduct a field experiment in Zambia's health assistant-training program. In their experiment,

⁴De Paola and Scoppa (2011) conduct a field experiment in an Italian university to prove that students taking midterm and final examinations obtain higher scores than students taking only final examinations, the difference being that the former receive information on the midterm performance.

⁵From a theoretical viewpoint, Becker and Rosen (1992) employ a rank-order tournament model to analyze student incentives to study an educational environment. Dubey and Geanakoplos (2010) compare absolute grading and relative grading, while Paredes (2016) constructs a theoretical model of relative grading employing an all-pay auction. In addition, Andreoni and Brownback (2017) undertake a laboratory experiment to relate an increase in the number of enrollments in an all-pay auction to a decrease in the uncertainty of the threshold needed to pass in a tournament.

student rewards are absolute, with some students advised they will receive a rank-related reward. The authors conclude that the performance of students whose initial achievement level is relatively low is significantly lower when told that they will receive a rank-related reward. In contrast, Jalava et al. (2015) conduct a field experiment in Swedish primary schools to compare schemes with several types of reward and find that student performance is higher when rewarded relatively. Czibor et al. (2014) conduct a field experiment in a Dutch university and compare relative and absolute grading. They find that motivated male students rewarded relatively tend to obtain higher performance than students rewarded absolutely. However, both female and male students are largely unaffected by the reward scheme. On this basis, Czibor et al. (2014) contend that rank incentives are weak if students adopt just-pass behavior. That is, if students only care about whether they can pass the course or not, they will not want to provide effort for any higher rank than that to which they aspired. Even if graded relatively, relative performance information feedback may exert different impacts on student incentives to study depending on their attitude toward getting higher grades.

As explained earlier, by considering students who care only about whether they can pass the course enables us to identify the impact of relative performance information feedback unrelative grading in a binary grading environment. As also discussed, most university students in Japan only care about whether they can pass the course. The School Basic Survey, which is conducted by the Ministry of Education, Culture, Sports, Science and Technology in Japan, shows that the university advancement rate was about 50% in 2017, whereas it was less than 20% in 1990. Increasing the university advancement rate in Japan could reflect demand-side forces that lead to greater incentives to pursue university education. Most Japanese firms distinguish between new college graduates and other job seekers, in the sense that many vacancies for regular full-time jobs explicitly target the former (Genda et al., 2010). Moreover, Lise et al. (2014) show that the wage premium for education, as measured by the ratio of mean wages for college-educated full-time workers to less-than-college-educated full-time workers, has tended to increase in Japan since 2000. Interestingly, while the admission process to prestigious Japanese universities is competitive, credit standards in college courses are low (Abe 2002). That is, hard work is not generally needed for graduation.⁶ The hiring policy of many Japanese

⁶Kaji (2015) notes that a nontrivial proportion of students in Japanese universities eschew classes that demand serious reading and homework, instead preferring classes in which they will easily earn a

firms also discourages students from studying hard to obtain better grades. For example, before 2013, most Japanese firms did not consider a student's grade point average (GPA) when seeking recruits in the labor market for new graduates.⁷ Taken together, the high college wage premium, the low credit standards in college courses, and recruitment policies in new graduate labor markets that ignore student GPAs, may result in students caring only about whether they can pass the course, and so they attempt to pass their courses with minimal effort cost.

This background contextualizes our research question. Does relative performance information feedback improve student examination scores under a relative grading scheme in a binary grading environment? To examine this, we conduct a field experiment to investigate relative performance information feedback in a relative grading scheme.

III Experimental Design

Description of a randomized trial This section provides the details of a randomized trial, which was performed using first-year students in an economics department at a Japanese private university. We begin by describing the flow of interventions in the experiments, which are displayed in Figure 1. The academic year comprised first and second semesters. The first semester began in April 2012 and ended in July 2012. The second semester began in September 2012 and ended in January 2013. We conducted a mathematical achievement test (referred to as the Pretest of Mathematics) immediately following university entrance. Students were enrolled in two compulsory introductory economics courses in their first year: Economics I in first semester and Economics II in second semester. In Economics I and II, we administered midterm and final examinations to grade students. While the midterm and final examinations in Economics I were in May and July 2012, those in Economics II were in November 2012 and January 2013. We note that the score for the Pretest of Mathematics was independent of the grades for Economics I and II. The dotted vertical lines in Figure 1 represent the timing of the examinations.

[Insert Figure 1 here]

credit (*raku-tan*) and seeking off-campus experiences.

⁷According to the *Nihon Keizai Shinbun* (in Japanese), the reason for this is that most firms considered the GPA an unreliable indicator of student attainment. Because each university has its own grading guidelines, firms also cannot account for the difference in GPA between one university and another (*Nihon Keizai Shinbun*, December 8, 2013).

We conducted our randomized control trial immediately after the midterm examination in Economics II and randomly assigned all students to the treatment or control group. Following the midterm examination, we gave letters to students revealing their score for the midterm examination, delivered to them in person during class time. In addition, the letters given to students in the treatment group also reported their ranks in the midterm examination. We did not include this information in the letters to the students in the control group. The student letter content is similar to that used by Ashraf et al. (2014). Figures 4 and 5 in Appendix C reproduce the information provided to the students in the treatment and control groups. On this basis, while students in the treatment group knew their precise rank, students in the control group would only have a vague awareness.

Because of the Pretest of Mathematics, we can focus only on those students who care about whether they can pass the course, although some students are also eager to get a better grade. We divided students into four classes according to their score in the Pretest of Mathematics. We regard students who mark higher in the Pretest of Mathematics as students who are willing to get a better grade as the score is independent of their grades. In particular, we placed all students with a top-40 score in the one small class. Hereafter, we refer to this as Classroom 1. We then randomly allocated the remaining students to the other three classes. Hereafter, we refer to these as Classrooms 2, 3, and 4. We regard students in Classroom 1 as those eager to get a better grade, but students in Classrooms 2, 3, and 4 as students who only care about whether they can pass the course. We fixed all class enrollments and instructors across both semesters. However, even though each class had its own instructors, all students took the same examination at the same time.

There are two points to note in our randomized control trial. First, some students did not receive the letter regarding the midterm examination because they were not in class. However, the grade is independent of course attendance, and we later confirmed the robustness of our experimental results with this in mind. Second, some students may have exchanged their rank information. Because our experimental design is similar to that of Tran and Zeckhauser (2012), except for the grading scheme, we share the problem that students in both the control and treatment groups sit in the same classroom, making this exchange of rankings a very real possibility. We discuss this further in Section V.3.

The grading scheme In our experiment, the uncertain threshold plays a crucial role in determining whether students pass the course. In this department, credits for both Economics I and II are required for graduation. Therefore, successfully completing the courses is a primary concern for all students. In both Economics I and II, instructors determine a threshold between pass and failure so that a pass rate is reasonable in the following sense. Officially, the university’s guidelines recommend that the standard “absolute” pass score for the course is 60 out of 100. However, if students were to be strictly graded using the guideline, then the number of students who fail the course would frequently be inordinately high. Because credits in both Economics I and II are mandatory, a student who fails the course must repeat the course the following year, and for whom there is a separate class. Because the capacity of the repeating class is limited and there is only one classroom, instructors have to adjust the pass scores to obtain “a reasonable” pass rate.⁸ For example, in Economics I, the instructors decided the pass rate was the top 87.3% percent of students, and so they set the cutting score at 51 rather than 60.

There are also more than 200 participants, and students compete not only with students in their own classroom but also the other classrooms. Moreover, the cutting score is common to the four classes. Therefore, whether students pass or fail will depend on their relative position in the score distribution of all students. However, students do not know the exact cutting score in advance of taking their examinations. Our experiment can reveal whether the difference in the precision of their relative performance information accounts for the difference in final examination scores under a relative grading setup. Note that students were already aware of this evaluation system. This is because they had already experienced it in Economics I, and the instructors repeated its details again at the beginning of the second semester.⁹

Balance between the control and treatment groups Table 1 provides the total number of students and the means and standard deviations of the midterm examination scores in Economics II for the control and treatment groups. Table 1 also shows how

⁸Instructors also determine the scores between one grade and another.

⁹In Economics II, the score calculation was as follows: “40% of the midterm examination score” + “60% of the final examination score” + “the number of homework submissions” (there were 10 homework assignments each worth one point). Perfect scores for both examinations were 110 points. From the total score of 110 points, 100 points were for the economics part, and the remaining 10 points for basic mathematics.

we randomly divided these students into the control and treatment groups. The total number of students who took the midterm examination amounts to 284 and the mean score is 49.57. We randomly divided these students into control and treatment groups. However, some students failed to receive the letter. Then, in our experiment, the number of subjects amounts to 255 and the mean score is 50.67. The number of students for the control and treatment groups are 130 and 125, respectively. The mean scores for the control and treatment groups are 51.48 and 49.82, respectively, and there is no significant difference in the mean scores between the control and treatment groups, as shown in row (a) in Panel B.

Although there were students who failed to receive the letter, they were also randomly divided into the treatment and treatment groups. The total number of students who failed to receive the letter amounted to 29 and the mean score is 39.90. The mean score of the students who received the letters is significantly different from that of the students who did not, as shown in row (b) in Panel B. However, the number of students who were assigned to the control group but did not receive the letter amounts to 13 and the mean score is 40.62, while the number of students who were assigned to the treatment group, but did not receive the letter amounts to 16 and the mean score is 39.31. We fail to reject the null hypothesis that “the mean values of these two groups are not different”, as shown in row (c) in Panel B. These results suggest that students who did not receive the letter were also randomly divided into the treatment and treatment groups.

[Insert Table 1 here]

Another point to note is the differences between classrooms. Table 1 also shows that we randomly divided students into the control and treatment groups if we take these differences into account. The mean score in the midterm examination in Classroom 1 is much higher than that in Classrooms 2–4, because students with a top-40 mark in the Pretest of Mathematics were enrolled in Classroom 1. The number of students who received the letter in Classrooms 2–4 is 215, while that in Classroom 1 is 40. The mean for students who received the letter in Classroom 1 is 65.48 and that in Classrooms 2–4 is 47.92. We reject the null hypothesis that “the mean values of the two groups are not different” in row (d) in Panel B. On the other hand, there is no significant difference in the mean scores between the control and treatment groups within Classrooms 2–4. While the number of students and the mean for the control group are 106 and 48.21,

respectively, those for the treatment group are 109 and 47.63, respectively. We do not reject the null hypothesis that “the mean values of the two groups are not different”, as shown in row (e) in Panel B. In addition, as for Classroom 1, the number of students for the control and treatment groups are 24 and 16, respectively, and the mean scores for the control and treatment groups are 65.96 and 64.75, respectively. We again do not reject the null hypothesis that “the mean values of the two groups are not different”, as shown in row (f) in Panel B.

IV A Theoretical Model

In this section, we construct a simple model in which students choose their effort input under uncertainty to pass the course. As the cutting score depends on the class score distribution in our experiment, strategic interaction among students may play a role in determining the effort input. However, as the number of students is sufficiently large, each student recognizes that a change in her own action makes little difference in the distribution of the scores of all students. We assume that any strategic interaction can be assumed away in the theoretical model. Nonetheless, this simple model helps us understand several key theoretical insights.

Setup Consider a course where the assessment comprises a midterm examination and a final examination. The course grade awarded for each student depends on the final score, which is the sum of the scores obtained in the two examinations: $s_0 + s$, where s_0 and s are the scores for the midterm and the final examinations, respectively.¹⁰ Before the course commences, the cutting score is set to $S \in \mathbb{R}$ and is known to all students. However, as previously described, students know that the actual cutting score can change according to the profile of all student final scores to realize a reasonable distribution of student grades. Therefore, S is not a rigorous criterion for course success, rather an approximate standard expected for the cutting score.

Introducing uncertainty in the cutting score, we model this environment as follows. A student with score s_0 in the midterm examination then believes that the average required

¹⁰We could consider the weighted-average score, $\phi s_0 + (1 - \phi)s$, according to the actual evaluation method, but the basic results do not change.

score in the final examination is equal to¹¹

$$\bar{s} \equiv S - s_0,$$

but the actual cutting score, denoted by \hat{s} , is a random variable of form

$$\hat{s} = \theta \bar{s}, \tag{1}$$

where θ is a random variable that follows a uniform distribution $[1 - \varepsilon, 1 + \varepsilon]$ and $\varepsilon \in (0, 1)$.

Following the midterm examination, every student in the class knows their own score $s_0 \in \mathbb{R}$. Each student then makes effort $e \geq 0$ in preparing for the final examination, and obtains a score

$$s = e \tag{2}$$

in the final examination, but at a cost of

$$c = \frac{e^2}{2}. \tag{3}$$

Finally, we assume that students are concerned only about whether they pass the course.¹² For any final examination scores s , the utility of the student is given by

$$u(s) = \begin{cases} 1 & \text{if } s \geq \hat{s}, \\ -1 & \text{if } s < \hat{s}. \end{cases}$$

Equilibrium The students select their level of effort e to maximize their expected utility. Mathematically, the optimization problem for students is

$$\begin{aligned} &\text{Maximize} && U(e) \equiv \mathbb{E}[u(s) - c], \\ &\text{subject to} && (2), (3), e \geq 0. \end{aligned}$$

¹¹When there is no uncertainty in the cutting score, the condition for passing the examination would be $s_0 + s \geq S$. This is equivalent to $s \geq \bar{s}$.

¹²The reward system in our experiment is a tournament in the sense students are evaluated relatively. However, because there are many participants and many winners, we assume that participants do not care about the performance of other participants directly, but rather pay attention to the borderline between winning and losing. Of course, some students will still attempt to obtain the best grade possible. However, to keep the model simple and to clearly understand the role of the borderline, we assume this preference relation. In the context of tournaments, Aoyagi (2010) and Ederer (2010) theoretically explore the relationship between information feedback and agent incentives.

Through simple calculations, we obtain the functional form of U ¹³:

$$U(e) = \begin{cases} -1 - \frac{e^2}{2} & \text{if } 0 \leq e < (1 - \varepsilon)\bar{s}, \\ -\frac{1}{\varepsilon} + \frac{1}{\varepsilon\bar{s}}e - \frac{e^2}{2} & \text{if } (1 - \varepsilon)\bar{s} \leq e < (1 + \varepsilon)\bar{s}, \\ 1 - \frac{e^2}{2} & \text{if } e \geq (1 + \varepsilon)\bar{s}. \end{cases}$$

We adopt the following assumption to focus on the solution most relevant to the empirical analysis.

Assumption. *Uncertainty for students is sufficiently strong:*

$$\varepsilon > \frac{1}{3}.$$

If ε is small, then each student selects either zero or minimal input for passing the examination regardless of s_0 . This is because students can deterministically control their grades because of weak uncertainty and hence weigh the certain benefits of qualification and the required effort costs. In contrast, if ε is sufficiently large, then students cannot do this when their scores in the midterm examination are of an intermediate value. Because the latter case is relevant for the experimental results, we adopt this assumption to exclude the former irrelevant case.

Proposition. *Let $s_0 = S - [2\varepsilon(1 - \varepsilon)]^{-\frac{1}{2}}$ and $\bar{s}_0 = S - [\varepsilon(1 + \varepsilon)]^{-\frac{1}{2}}$. In equilibrium, the optimum e^* is characterized by*

$$e^* = \begin{cases} 0 & \text{if } s_0 < \underline{s}_0, \\ \frac{1}{\varepsilon\bar{s}} & \text{if } \underline{s}_0 \leq s_0 < \bar{s}_0, \\ (1 + \varepsilon)\bar{s} & \text{if } s_0 \geq \bar{s}_0. \end{cases}$$

Proof. See Appendix B.

The intuition underlying the proposition is as follows. If a student obtains a low score in the midterm examination (that is, $s_0 < \underline{s}_0$), they will make no effort because an unacceptably large effort is required in the final examination to pass the course. In contrast, if their grade in the midterm examination is sufficiently high (that is, $s_0 \geq \bar{s}_0$), the student will minimize their effort under the constraint of a cutting score because this

¹³The derivation is in Appendix A.

can be at a very small effort cost. Finally, in the intermediate case (that is, $\underline{s}_0 \leq s_0 < \bar{s}_0$), the student selects the inner solution as a means of balancing the uncertain benefit and certain cost. In this case, the student cannot predict with certainty the result of the final examination. Hereafter, we say students belong to: *the lower class* if their scores in the midterm examination scores lie within the range of $s_0 < \underline{s}_0$; *the middle class* if their scores in the midterm examination scores lie within the range of $\underline{s}_0 \leq s_0 < \bar{s}_0$; and *the upper class* if their scores in the midterm examination scores lie within the range of $\bar{s}_0 \leq s_0$.

As \bar{s} is decreasing in s_0 , we find that e^* is increasing in $[s_0, \bar{s}_0)$ and decreasing in $(\bar{s}_0, +\infty)$ with respect to s_0 . Thus, e^* has a single peak at \bar{s}_0 . Figure 2 provides numerical examples of the relationship between s_0 and $s^* = e^*$, the equilibrium score in the final examination.

[Insert Figure 2 here]

Empirical Implications We focus on the relationship between grading uncertainty and the final examination scores. A change in uncertainty has two effects on the equilibrium final examination score. First, given the classification, information provision changes the equilibrium score: $\frac{1}{\varepsilon \bar{s}}$ and $(1 + \varepsilon)\bar{s}$. Second, a change in uncertainty affects the thresholds of classification, \underline{s}_0 and \bar{s}_0 . In particular, we focus on the upper threshold, \bar{s}_0 . We obtain the following immediately from the proposition.

Result 1 *A decrease in the degree of uncertainty improves the final examination scores of students in the middle class ($\underline{s}_0 \leq s_0 < \bar{s}_0$). However, this exerts a detrimental effect on the final examination scores for students in the upper class ($s_0 \geq \bar{s}_0$).*

Result 2 *A decrease in the degree of uncertainty lowers the upper threshold, \bar{s}_0 .*

Result 1 demonstrates that the effects of information feedback on the final examination scores vary according to the range of midterm examination scores. In particular, the information feedback has positive (negative) effects for students in *the middle (upper) class*. Information feedback thus encourages students in *the middle class*. This is because it reduces the risk of a fail in the final examination and therefore induces them to increase their required effort to achieve a cutting score. However, this also expands the

opportunity for students in *the upper class* to economize on the required level of effort to pass the final examination. Result 2 shows that information feedback converts some *middle class* students into *upper class* students. Intuitively, the reason is that information feedback notifies students near the (old) upper threshold so that they can pass the exam with certainty at small cost.

A consequence of information feedback is theoretically ambiguous around the upper threshold. While Figure 2 provides an example in which the total effect is complicated, we evaluate the total effect using the actual experiment.

V Estimation Model and Results

V.1 Estimation Model

From a theoretical point of view, the ε of students in the treatment group is lower than the ε of students in the control group. This is because students in the treatment group have information about both their own midterm scores and rank, whereas students in the control group have information only on their own midterm scores.

According to the proposition and Figure 2, the final score increases as the midterm score increases (decreases) for students whose midterm scores are intermediate (high). A quadratic specification enables us to capture this relationship. In addition, according to Result 1, for students whose midterm scores are sufficiently high, the students in the control group tend to achieve higher scores than those in the treatment group. In contrast, for students whose midterm scores are intermediate, the students in the control group tend to achieve lower scores than those in the treatment group.

We employ the following empirical framework:

$$Y_{Fi} = \alpha \left(\frac{Y_{Mi}^2}{100} \times D_i \right) + \beta (Y_{Mi} \times D_i) + \gamma D_i + \theta \frac{Y_{Mi}^2}{100} + \phi Y_{Mi} + \mathbf{X}_i \boldsymbol{\delta} + \epsilon_i, \quad (4)$$

where Y_{Fi} and Y_{Mi} denote the respective scores in the final and the midterm examinations for student i , $\frac{Y_{Mi}^2}{100}$ is the squared midterm score for student i , divided by 100. D_i is a dummy variable equal to one if student i is given information on her relative rank in the midterm examination (i.e., the student is in the treatment group), and zero if student i is not given this information (i.e., the student is in the control group). $\frac{Y_{Mi}^2}{100} \times D_i$ and

$Y_{Mi} \times D_i$ are interaction terms for both variables.¹⁴ The vector \mathbf{X}_i contains the student characteristics, including the number of homework submissions, dummy variables for male students and students in different classrooms, and a constant term. ϵ_i are disturbances, which we assume are distributed $N(0, \sigma^2)$. The definitions of all variables in equation (4) are in Table 2.

[Insert Table 2 here]

The coefficients upon which we focus are α and β , which are the estimated coefficients for the interaction terms. These coefficients enable us to determine whether the impact of information on a student’s relative rank in the midterm examination and the improvement (if any) in the final examination score vary in accordance with the midterm examination score.¹⁵

V.2 Descriptive Statistics

In the baseline estimation model, the sample is restricted to students who meet the following criteria: (i) students who received the letter, (ii) students in Classrooms 2–4, and (iii) students for whom information on all the relevant variables is available.¹⁶ Table 2 displays descriptive statistics for all the variables used in the baseline estimation model. Comparing the mean scores in the midterm examination Y_M for the control and treatment groups, the means (standard deviations) for the control and treatment groups are 48.21 (17.70) and 47.44 (13.82), respectively. We fail to reject the null hypothesis that “the mean values of the two groups are not different”.

To visualize the relationship between the midterm and final examination scores, we provide a scatterplot for the control and treatment groups in Figure 3. We can see that improvements in the final examination scores depended on the range of midterm examination scores. These observations suggest that the impact of information on a student’s relative rank in the midterm examination on the extent of improvement in the final examination varies according to the midterm examination score.

¹⁴As Barnett et al. (2005) argue, randomized experiments can reduce the effect of the regression to the mean (RTM). As the responses from both the control and treatment groups are equally affected by the RTM, the differences between the treatment group and the control group, that is, the coefficients for D_i and the interaction terms, comprise the treatment effect after adjusting for the RTM.

¹⁵All regression results in this section are estimated using Stata version 14.

¹⁶We exclude three students who took a makeup examination, whose midterm score was revised, or who made a mistake in writing on their answer sheet from the sample.

[Insert Figure 3 here]

V.3 Estimation Results

V.3.1 The quadratic specifications

The estimation results in Table 3 demonstrate that the final score is an inverted U-shaped function of the mid score, which is consistent with the proposition. The estimation results for the baseline model (equation (4)) are in column (1). The coefficient for $\frac{Y_M^2}{100}$ is significantly negative ($\theta = -0.758$) and the coefficient for Y_M is significantly positive ($\phi = 1.546$). These values indicate that for students in the control group, the final score is an inverted U-shaped function of the mid score. Holding everything else constant, it is easy to see that the value of the final scores are maximized, that is, the upper threshold, when $Y_M^* = \frac{-50\phi}{\theta}$ for the control group and $Y_M^* = \frac{-50(\beta+\phi)}{\alpha+\theta}$ for the treatment group, respectively. Because the value of the upper threshold Y_M^* for the control group is 101.98, *the upper class* cannot be observed in the control group. That is, students who receive a higher midterm score tend to get a higher final score regardless of their midterm score. On the other hand, the coefficient for $\frac{Y_M^2}{100} \times D$ is significantly negative ($\alpha = -1.457$) and the coefficient for $Y_M \times D$ is positive ($\beta = 1.113$). The value of the upper threshold Y_M^* for the treatment group is 60.03. These values indicate that for the students in the treatment group, the final score is also an inverted U-shaped function of the mid score. The students whose midterm score is lower (higher) than 60 in the treatment group are theoretically classified as *the middle (upper) class*.¹⁷

[Insert Table 3 here]

We calculate the fitted values of the final score, categorized by the control and treatment groups, using the estimated coefficients of the variables in column (1) of Table 3. As shown in Figure 3, we clearly see that for students in *the middle class* (who scored under approximately 60), the more precise the relative performance information provided, the higher the score in the final examination. In contrast, for students with relatively high scores (over approximately 60) in the midterm examination, those in the treatment group obtained lower scores in the final examination than those in the control group.

¹⁷When we exclude students whose scores in the midterm examination are more than two standard deviations from the mean from our analysis as potential outliers, the estimation results are almost identical to those shown in column (1).

The baseline result shown in column (1) is robust even if we consider potential self-selection bias. As discussed in Section III, some students did not receive the letter. These students were excluded from the baseline estimation result reported in column (1). However, as some students decided not to receive the letter (by not attending class, possibly because they were unwilling to study hard), there is a potential for self-selection bias. To confirm the robustness of the estimation result in column (1), we added students who did not receive the letter into the sample. U_i denotes a dummy variable equal to one if student i did not receive the letter and zero if student i received the letter. $\frac{Y_{Mi}^2}{100} \times U_i$ and $Y_{Mi} \times U_i$ are interaction terms, respectively. As shown in column (2), the coefficient for $\frac{Y_M^2}{100}$ is significantly negative and the coefficient for Y_M is significantly positive. On the other hand, the coefficient for $\frac{Y_M^2}{100} \times D$ is significantly negative and the coefficient for $Y_M \times D$ is positive. These values are quite close to those reported in column (1). These indicate that our baseline result is quite robust, even if we include students who did not receive the letter.

Another question is whether the baseline result is robust when some students are eager to receive a better grade. While most students care only about whether they can pass the course, some students are keen to earn the highest possible grade. As discussed in Section III, we placed all students with a top-40 score in the Pretest of Mathematics in Classroom 1. We consider these students as being motivated compared with students in Classrooms 2, 3, and 4. For this reason, we excluded them from our baseline sample identifying the impact of relative performance information feedback on the incentives of students who only care about whether they can pass the course. To confirm the robustness of our baseline result, we then incorporated the students in Classroom 1 into the sample.¹⁸ As reported in column (3), the coefficient for $\frac{Y_M^2}{100}$ is significantly negative (-0.762) and the coefficient for Y_M is significantly positive (1.561). Conversely, the coefficient for $\frac{Y_M^2}{100} \times D$ is significantly negative (-1.441) and the coefficient for $Y_M \times D$ is positive (1.098). These values are quite close to those reported in column (1). Our baseline result is then quite robust, even when we consider the presence of motivated students. The estimation results reported in column (4) reflect where we added students who did not receive the letter

¹⁸We additionally included the following variables into the baseline model (equation (4)); $Class1_i$, $\frac{Y_{Mi}^2}{100} \times Class1_i$, $Y_{Mi} \times Class1_i$, $\frac{Y_{Mi}^2}{100} \times D_i \times Class1_i$, $Y_{Mi} \times D_i \times Class1_i$ and $D_i \times Class1_i$. We exclude students whose scores in the midterm examination exceed two standard deviations from the mean as potential outliers in the mathematics class.

into the sample reported in column (3). The estimated coefficients for $\frac{Y_M^2}{100}$, Y_M , $\frac{Y_M^2}{100} \times D$, and $Y_M \times D$ in column (4) are also quite close to those reported in column (1).¹⁹

In terms of other research considerations, such as the experimental design employed by Tran and Zeckhauser (2012), we divided students into control and treatment groups within each classroom. Because of this, we cannot discount the possibility that students shared their ranks with their classmates because we are unable to restrict student interactions outside the classroom entirely. However, it would be generally difficult for a student in the control group to identify a student in the treatment group with exactly the same score.²⁰ In addition, when students exchange information to determine their own rank, this leads to an underestimation (toward zero) in the estimated treatment effects. If so, our estimated treatment effects—that for students with intermediate (high) midterm examination scores, the more precise the information provided, the higher (lower) their score in the final examination—might be smaller than the true treatment effects.

V.3.2 The heterogeneous impacts due to the students' midterm scores

Based on the results in Table 3, Table 4 investigates the empirical implications from Results 1 and 2, which argue that the relative performance information feedback has heterogeneous impacts due to the students' midterm scores. Result 1 points out that the relative performance information feedback raises effort for students in *the middle class* but reduces the effort for students in *the upper class*. However, Result 2 demonstrates that the threshold between *the upper* and *middle classes* decreases because of a decrease in uncertainty. That is, there is a range of midterm scores where students who know their ranks are in *the upper class* but students who do not are in *the middle class*. Therefore, the relative performance information feedback has ambiguous effects on these students. According to the results in Table 3, when the range of midterm scores is lower than 60 (the value of the upper threshold for the treatment group), students in both the control and treatment groups belong to *the middle class*. For students whose midterm scores are below 60, the relative performance information feedback would then have a positive impact on their performance. In contrast, when the range of midterm scores is over 60,

¹⁹One point to note is that the relative performance information feedback has an insignificant impact on the students in Classroom 1. The estimation results are available upon request.

²⁰Further, we also perceive that many Japanese students do not like revealing their grades to friends. As observed by Benedict (1946), Japan has a long tradition as a shame society where students receiving higher (lower) grades prefer to avoid the envy (derision) of those with lower (higher) grades.

students in the treatment group belong to *the upper class*, while students in the control group still belong to *the middle class*. For students whose midterm scores are over 60, the impact of relative performance information feedback on their performance would be ambiguous.

In order to identify the heterogeneous impact of the relative performance information feedback, we consider the following equation:

$$Y_{Fi} = \gamma D_i + \phi Y_{Mi} + \lambda H_i + \beta(D_i \times H_i) + \zeta(Y_{Mi} \times H_i) + \mathbf{X}_i \boldsymbol{\delta} + \epsilon_i, \quad (5)$$

where H_i is a dummy variable equal to one if student i 's midterm score is higher than 60. Table 4 demonstrates the significance in the relative performance information feedback. As reported in column (1), the coefficient for D is significantly positive ($\gamma = 5.066$). This indicates that the relative performance information feedback has a substantial positive impact on the performance of students whose midterm scores are below 60. In contrast, the coefficient for the interaction terms, $D \times H$, is significantly negative ($\beta = -12.190$), and the coefficient for $D + (D \times H)$ is also significantly negative (f-value = 5.90). That is, for students whose midterm scores are over 60, the relative performance information feedback has a negative impact on their performance. In other words, the relative performance information feedback only exploits the incentive to study of students in *the middle class*. By way of contrast, it makes students whose midterm scores are relatively high slacken off.

[Insert Table 4 here]

These empirical results are robust, even if we consider the potential self-selection bias and the presence of students in Classroom 1. As for the potential self-selection bias, we include U_i , $Y_{Mi} \times U_i$, $H_i \times U_i$, and $Y_{Mi} \times H_i \times U_i$ into the equation (5). As shown in column (2), the coefficient for D is significantly positive ($\gamma = 5.213$). In contrast, the coefficient for the interaction terms, $D \times H$, is significantly negative ($\beta = -12.104$), and the coefficient for $D + (D \times H)$ is also significantly negative (f-value = 5.51). These values are quite close to those reported in column (1). These indicate that our baseline result is quite robust, even if we consider the students who did not receive the letter. As for the presence of students in Classroom 1, as shown in column (3), the coefficient for D is significantly positive ($\gamma = 5.061$).²¹ In contrast, the coefficient for the interaction terms,

²¹The model shown in column (3) includes $Class1_i$, $D_i \times Class1_i$, $Y_{Mi} \times Class1_i$, $H_i \times Class1_i$, $D_i \times H_i \times Class1_i$ and $Y_{Mi} \times H_i \times Class1_i$ into the equation (5).

$D \times H$, is significantly negative ($\beta = -12.120$), and the coefficient for $D + (D \times H)$ is also significantly negative (f-value = 5.76). These values are quite close to those reported in column (1). Our result is then quite robust even when we consider the presence of students in Classroom 1. The estimation results reported in column (4) are from the case where we considered both the potential self-selection bias and the presence of students in Classroom 1. The estimated coefficients for D , Y_M and $D \times H$ in column (4) are also quite close to those reported in column (1).

VI Concluding Remarks

Our experimental results demonstrate that relative performance information feedback on final scores has opposing effects depending on a student's midterm score. For students with average scores in the midterm examination, relative performance information serves as a signal indicating that they will need to work harder to pass the course. In contrast, for students with sufficiently high scores in the midterm examination, relative performance information is a signal that they can pass with only moderate effort.

Our results suggest that the information on relative performance is not necessarily beneficial in incentivizing students to study in the binary grade environments. For students with intermediate midterm examination scores, it is beneficial to inform them of their performance ranking as a means of decreasing uncertainty. In contrast, for students with high midterm examination scores, providing their own relative performance information could be detrimental to their overall performance in the course. From an examiner or policy maker perspective, the relative performance information is favorable if examiners prefer encouraging students who lie on the threshold. This is not the case if examiners prefer exploiting the effort of already high-performing students.

We conclude the paper by discussing the prediction for an education environment where there are multiple grades and students are eager to earn their highest possible grade. Our findings suggest that relative performance information feedback may also have a positive impact on the incentive to study here. In multiple threshold environments, a student above the threshold between B and C is also below the threshold between A and B. If students prefer to receive their highest possible grade, the positive aspect of the relative performance feedback always dominates its negative aspect. In this case, relative performance information feedback would always be favorable. It remains to

future research to investigate these predictions.

Appendix A Derivation of the Functional Form of $U(e)$

For the case of $s = e < (1 - \varepsilon)\bar{s}$, students fail the examination for certain. Therefore, $U(e) = -1 - \frac{e^2}{2}$. At the same time, $s = e \geq (1 - \varepsilon)\bar{s}$ ensures that the students pass the examination, that is, $U(e) = 1 - \frac{e^2}{2}$. If $(1 - \varepsilon)\bar{s} \leq s = e < (1 + \varepsilon)\bar{s}$, whether students pass or fail the examination depends on the realized \hat{s} . Because $\hat{s} \leq s = e \Leftrightarrow \theta \leq \frac{e}{\bar{s}}$, $U(e)$ can be written as follows:

$$\begin{aligned} U(e) &= \int_{1-\varepsilon}^{\frac{e}{\bar{s}}} 1 \times \frac{1}{2\varepsilon} d\theta + \int_{\frac{e}{\bar{s}}}^{1+\varepsilon} (-1) \times \frac{1}{2\varepsilon} d\theta - \frac{e^2}{2} \\ &= -\frac{1}{\varepsilon} + \frac{1}{\varepsilon\bar{s}}e - \frac{e^2}{2}. \end{aligned}$$

Appendix B Proof of Proposition

Put $\tilde{U}(e) = -\frac{1}{\varepsilon} + \frac{1}{\varepsilon\bar{s}}e - \frac{e^2}{2}$. Let \tilde{e} be the unique maximizer of \tilde{U} , $\frac{1}{\varepsilon\bar{s}}$. As U is decreasing in $[0, (1 - \varepsilon)\bar{s})$ and $[(1 + \varepsilon)\bar{s}, +\infty)$, U can be maximized only at 0, \tilde{e} or $(1 + \varepsilon)\bar{s}$.

Case 1: $\tilde{e} < (1 - \varepsilon)\bar{s}$

In this case, $e^* = 0$ because U is decreasing throughout the domain. We obtain that $\tilde{e} < (1 - \varepsilon)\bar{s}$ if and only if

$$s_0 < S - [\varepsilon(1 - \varepsilon)]^{-\frac{1}{2}} \equiv s_0^1.$$

Case 2: $(1 - \varepsilon)\bar{s} \leq \tilde{e} < (1 + \varepsilon)\bar{s}$

The condition $(1 - \varepsilon)\bar{s} \leq \tilde{e} < (1 + \varepsilon)\bar{s}$ is equivalent to

$$s_0^1 \leq s_0 < S - [\varepsilon(1 + \varepsilon)]^{-\frac{1}{2}} \equiv s_0^2.$$

In this case, $e^* = 0$ or \tilde{e} because U is decreasing for $e \geq \tilde{e}$. The condition for $e^* = \tilde{e}$ is $U(0) \leq U(\tilde{e})$, which can be reduced to

$$s_0 \geq S - [2\varepsilon(1 - \varepsilon)]^{-\frac{1}{2}} \equiv s_0^3.$$

It is obvious that $s_0^1 < s_0^3$. Besides, $s_0^3 < s_0^2$ under Assumption 1.

Therefore, $e^* = 0$ if $s_0 < s_0^3$ and $e^* = \tilde{e}$ if $s_0^3 \leq s_0 < s_0^2$.

Case 3: $(1 + \varepsilon)\bar{s} \leq \tilde{e}$

By Case 2, $(1 + \varepsilon)\bar{s} \leq \tilde{e}$ if and only if $s_0 \geq s_0^2$. In this case, $e^* = 0$ or $(1 + \varepsilon)\bar{s}$ because U is increasing in $[(1 - \varepsilon)\bar{s}, (1 + \varepsilon)\bar{s}]$. The condition for $e^* = (1 + \varepsilon)\bar{s}$ is $U((1 + \varepsilon)\bar{s}) \geq U(0)$ and it is equivalent to

$$s_0 \geq S - 2(1 + \varepsilon)^{-1} \equiv s_0^4.$$

By long but straightforward calculations, we find that $s_0^4 < s_0^2$ under Assumption 1. Hence, $e^* = (1 + \varepsilon)\bar{s}$ if $s_0 \geq s_0^2$.

In sum, we obtain the characterization of e^* in Proposition 1 by putting $s_0^3 = \underline{s}_0$ and $s_0^2 = \bar{s}_0$ (Q.E.D.)

Appendix C The letters to the students

[Insert Figures 4 and 5 here]

Acknowledgements

We are very grateful to Akira Yamazaki, Kentaro Kobayashi, Hayato Nakata, and Masahiro Watabe for invaluable advice. We also thank Naohito Abe, Kosuke Aoki, David Gill, Tao Gu, Shigeki Kano, Vu Tuan Kai, Jun-Hyung Ko, Akira Miyaoka, Tomoharu Mori, Chang Min Lee, David Reiley, Masao Nagatsuka, Kengo Nutahara, Daniela Puzzello, Kan Takeuchi, and participants at annual meeting on Japanese Economic Association 2013 Spring Meeting (Toyama, Japan), Economic Science Association European meeting 2015 (Heidelberg, Germany), and seminar participants at Osaka University, Seikei University, and the University of Tokyo for their helpful comments. This experiment was approved by the university's research ethics committee (Application No. H26-002). We have not received financial support for this research from any outside agency or foundation.

References

- Abe, Y. (2002). Universities and the entry-level job market: Evidence from Japanese panel data. *Labour Economics* 9(6): 699–715.
- Andreoni, J., and Brownback, A. (2017). All pay auction and group size: Grading on a curve and other applications. *Journal of Economic Behavior and Organization*, 137: 361-373.
- Aoyagi, M. (2010). Information feedback in a dynamic tournament. *Games and Economic Behavior* 70(2): 242–260.
- Ashraf, N., Bandiera, O., and Lee, S. S. (2014). Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior and Organization* 100: 44–63.
- Azmat, G., and Iriberry, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics* 94(7): 435–452.
- Azmat, G., and Iriberry, N. (2016). The provision of relative performance feedback: An analysis of performance and satisfaction. *Journal of Economics and Management Strategy* 25(1): 77–110.

- Bandiera, O., Larcinese, V., and Rasul, I. (2015). Blissful ignorance? A natural experiment on the effect of feedback on students' performance. *Labour Economics* 34: 13–25.
- Barnett, A. G., Van Der Pols, J. C., and Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology* 34(1): 215–220.
- Becker, W. E., and Rosen, S. (1992). The learning effect of assessment and evaluation in high school. *Economics of Education Review* 11(2): 107–118.
- Benedict, R. (1946) *The Chrysanthemum and the Sword: Patterns of Japanese Culture*. Mariner Books.
- Czibor, E., Onderstal, S., Sloof, R., and Van Praag, M. (2014). Does relative grading help male students? Evidence from a field experiment in the classroom. *IZA Discussion Paper*, No. 8429.
- De Paola, M., and Scoppa, V. (2011). Frequency of examinations and student achievement in a randomized experiment. *Economics of Education Review* 30(6): 1416–1429.
- Dubey, P., and Geanakoplos, J. (2010). Grading exams: 100, 99, 98, ... or A, B, C? *Games and Economic Behavior* 69(1): 72–94.
- Ederer, F. (2010). Feedback and motivation in dynamic tournaments. *Journal of Economics and Management Strategy* 19(3): 733–769.
- Eriksson, T., Poulsen, A., and Villeval, M. C. (2009). Feedback and incentives: Experimental evidence. *Labour Economics* 16(6): 679–688.
- Freeman, R. B., and Gelber, A. M. (2010). Prize structure and information in tournaments: Experimental evidence. *American Economic Journal: Applied Economics* 2(1): 149–164.
- Genda, Y., Kondo, A., and Ohta, S. (2010). Long-term effects of a recession at labor market entry in Japan and the United States. *Journal of Human Resources* 45(1): 157–196.
- Gill, D., Z. Kissovav, J. Lee, and V. Prowse (2018) "First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision." *Management Science* Published Online: <https://doi.org/10.1287/mnsc.2017.2907>

- Grant, A. (2016). Why we should stop grading students on a curve. *The New York Times*, accessed January 6, 2017. Available from URL: <https://www.nytimes.com/2016/09/11/opinion/sunday/why-we-should-stop-grading-students-on-a-curve.html?smid=tw-share&r=0>
- Jalava, N., Joensen, J. S., and Pellas, E. (2015). Grades and rank: Impacts of non-financial incentives on test performance. *Journal of Economic Behavior and Organization* 115: 161–196.
- Kaji, S. (2015). Japanese universities need brighter ideas. *Financial Times*, accessed March 6, 2017. Available from URL: <https://www.ft.com/content/bb4cb810-7717-11e5-a95a-27d368e1ddf7>
- Lazear, E. P., and Rosen, S. (1981). Rank-order tournament as optimum labor contracts. *Journal of Political Economy* 89(5): 841–864.
- Lise, J., Sudo, N., Suzuki, M., Yamada, K., and Yamada, T. (2014). Wage, income and consumption inequality in Japan, 1981–2008: From boom to lost decades. *Review of Economic Dynamics* 17: 582–612.
- Ludwig, S., and Lünser, G. K. (2012). Observing your competitor: The role of effort information in two-stage tournaments. *Journal of Economic Psychology* 33: 166–182.
- Massachusetts Court System (2017) “Massachusetts Bar Examination Grading Information,” accessed August 30, 2017. Available from URL: <http://www.mass.gov/courts/docs/bbe/k-website-2015-bar-exam-scoring-information.pdf>
- Paredes, V. (2016). Grading system and student effort. *Education, Finance and Policy*. 12(1): 107–128.
- Stenson, T. (2017) “Demystifying the New York and Massachusetts bar exams.” *BARBRI BAR review*, accessed August 30, 2017. Available from URL: https://hls.harvard.edu/content/uploads/2017/02/Barbri-NY-MA-Bar-Exams_February-2016-Stenson.pdf
- Tran, A., and Zeckhauser, R. (2012). Rank as an inherent incentive: Evidence from a field experiment. *Journal of Public Economics* 96(9): 645–650.
- Volokh, E. (2015). In praise of grading on a curve. *The Washington Post*, accessed March 6, 2017. Available from URL:

https://www.washingtonpost.com/news/volokh-conspiracy/wp/2015/02/09/in-praise-of-grading-on-a-curve/?utm_term=.63bfc5beb916

Table 1: Confirmation of randomness

Panel A. Descriptive statistics of the midterm examination scores				
		1	2	3
		Classrooms 1–4	Classrooms 2–4	Classroom 1
I. All	Obs.	284	244	40
	Mean	49.57	46.96	65.48
	S.D.	17.36	16.24	15.54
II. Receive a letter	Obs.	255	215	40
	Mean	50.67	47.92	65.48
	S.D.	17.02	15.85	15.54
-i. Control	Obs.	130	106	24
	Mean	51.48	48.21	65.96
	S.D.	18.64	17.70	15.84
-ii. Treatment	Obs.	125	109	16
	Mean	49.82	47.63	64.75
	S.D.	15.18	13.90	15.56
III. Not receive a letter	Obs.	29	29	
	Mean	39.90	39.90	
	S.D.	17.56	17.56	
-i. Control	Obs.	13	13	
	Mean	40.62	40.62	
	S.D.	16.75	16.75	
-ii. Treatment	Obs.	16	16	
	Mean	39.31	39.31	
	S.D.	18.73	18.73	

Panel B. Mean-comparison test (Welch t-test)			
		t-value	P-value
(a)	Comparison (1, II-i) with (1, II-ii)	0.781	0.435
(b)	Comparison (1, II) with (1, III)	3.140	0.003
(c)	Comparison (1, III-i) with (1, III-ii)	0.198	0.845
(d)	Comparison (2, II) with (3, II)	6.452	0.000
(e)	Comparison (2, II-i) with (2, II-ii)	0.264	0.792
(f)	Comparison (3, II-i) with (3, II-ii)	0.239	0.813

Table 2: Descriptive statistics

	Variables	Definition	Mean	Std. Dev.	Min	Max
Total n=214	Y_F	Score in the final examination	61.73	20.47	0	100
	Y_M	Score in the midterm examination	47.82	15.83	12	89
	D	=1 if student is given information on her relative rank in the midterm examination, =0 if elsewhere	0.50	0.50	0	1
	$Class1$	=1 if in the classroom A (math class), =0 elsewhere	0.00	0.00	0	0
	$Class2$	=1 if in the classroom B, =0 elsewhere	0.35	0.48	0	1
	$Class3$	=1 if in the classroom C, =0 elsewhere	0.31	0.46	0	1
	$Male$	=1 if male, =0 if female	0.87	0.34	0	1
	$Honework$	The number of homework submissions	5.25	3.62	0	10
Treatment n=108	Y_F		63.35	17.71	0	92
	Y_M		47.44	13.82	12	77
	D		1.00	0.00	1	1
	$Class1$		0.00	0.00	0	0
	$Class2$		0.35	0.48	0	1
	$Class3$		0.31	0.46	0	1
	$Male$		0.87	0.34	0	1
	$Honework$		5.25	3.61	0	10
Control n=106	Y_F		60.08	22.91	0	100
	Y_M		48.21	17.70	18	89
	D		0.00	0.00	0	0
	$Class1$		0.00	0.00	0	0
	$Class2$		0.35	0.48	0	1
	$Class3$		0.32	0.47	0	1
	$Male$		0.87	0.34	0	1
	$Honework$		5.25	3.65	0	10

Mean-comparison test (Welch t-test)

H_0 : there is no difference in the means of Y_M ; t-value=0.35 (p-value=0.73)

Table 3: Estimation results: the quadratic specifications

	Classrooms 2–4		Classrooms 1–4	
	Students did not receive the letter			
	Excluded	Included	Excluded	Included
	(1)	(2)	(3)	(4)
$\frac{Y_M^2}{100} \times D$	-1.457** [0.703]	-1.433** [0.703]	-1.441** [0.713]	-1.418** [0.712]
$Y_M \times D$	1.113 [0.733]	1.079 [0.731]	1.098 [0.743]	1.065 [0.740]
D	-14.313 [18.783]	-13.301 [18.702]	-13.991 [19.026]	-12.987 [18.913]
$\frac{Y_M^2}{100}$	-0.758* [0.443]	-0.786* [0.444]	-0.762* [0.448]	-0.790* [0.448]
Y_M	1.546*** [0.510]	1.566*** [0.511]	1.561*** [0.516]	1.581*** [0.516]
$\frac{Y_M^2}{100} \times U$		3.268* [1.756]		3.227* [1.785]
$Y_M \times U$		-2.695* [1.437]		-2.675* [1.456]
U		17.767 [29.950]		17.309 [30.296]
Observations	214	242	253	281
Log likelihood	-890.6	-1037	-1038	-1188
R^2	0.42	0.49	0.46	0.52
F-test H_0 : All coefficients except the cons. = 0	15.51***	28.14***	23.11***	32.48***

Notes

1) Standard errors in bracket are heteroskedasticity robust.

2) *, ** and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

3) The models in columns (1)–(4) also include the variables; *Class2*, *Class3*, *Males*, *Homework* and *Constant*. Estimates for these variables not reported.4) The models in columns (3) and (4) include the variables; *Class1*, $D \times Class1$, $Y_M \times Class1$, $H \times Class1$, $D_i \times H \times Class1$ and $Y_M \times H \times Class1$. Estimates for these variables not reported.

Table 4: Estimation results: the heterogeneous impacts of information feedback

	Classrooms 2–4		Classrooms 1–4	
	Students did not receive the letter			
	Excluded	Included	Excluded	Included
	(1)	(2)	(3)	(4)
D	5.066*	5.213*	5.061*	5.225*
	[2.756]	[2.795]	[2.778]	[2.810]
Y_M	0.903***	0.909***	0.915***	0.920***
	[0.134]	[0.138]	[0.136]	[0.139]
$D \times H$	-12.190***	-12.104***	-12.120***	-12.110***
	[4.095]	[4.120]	[4.108]	[4.132]
$Y_M \times H$	-0.655***	-0.744***	-0.653***	-0.743***
	[0.226]	[0.226]	[0.228]	[0.229]
H	41.918***	47.647***	41.748***	47.627***
	[14.710]	[14.687]	[14.820]	[14.867]
Observations	214	242	253	281
Log likelihood	-893.1	-1039	-1040	-1190
R^2	0.41	0.48	0.45	0.51
F-test H_0 :				
The coefficient for				
$D + (D \times H) = 0$	5.90**	5.51**	5.76**	5.45**
All coefficients except				
the cons. = 0	13.34***	18.26***	14.90***	16.16***

Notes

- 1) Standard errors in bracket are heteroskedasticity robust.
- 2) *, ** and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.
- 3) The models in columns (1)–(4) also include the variables; *Class2*, *Class3*, *Males*, *Homework* and *Constant*. Estimates for these variables not reported.
- 4) The models in columns (2) and (4) include the variables; U , $Y_M \times U$, $H_i \times U$, and $Y_M \times H \times U$. Estimates for these variables not reported.
- 5) The models in columns (3) and (4) include the variables; *Class1*, $D \times Class1$, $Y_M \times Class1$, $H \times Class1$, $D_i \times H \times Class1$ and $Y_M \times H \times Class1$. Estimates for these variables not reported.

Figure 1: The flow of interventions in the experiment

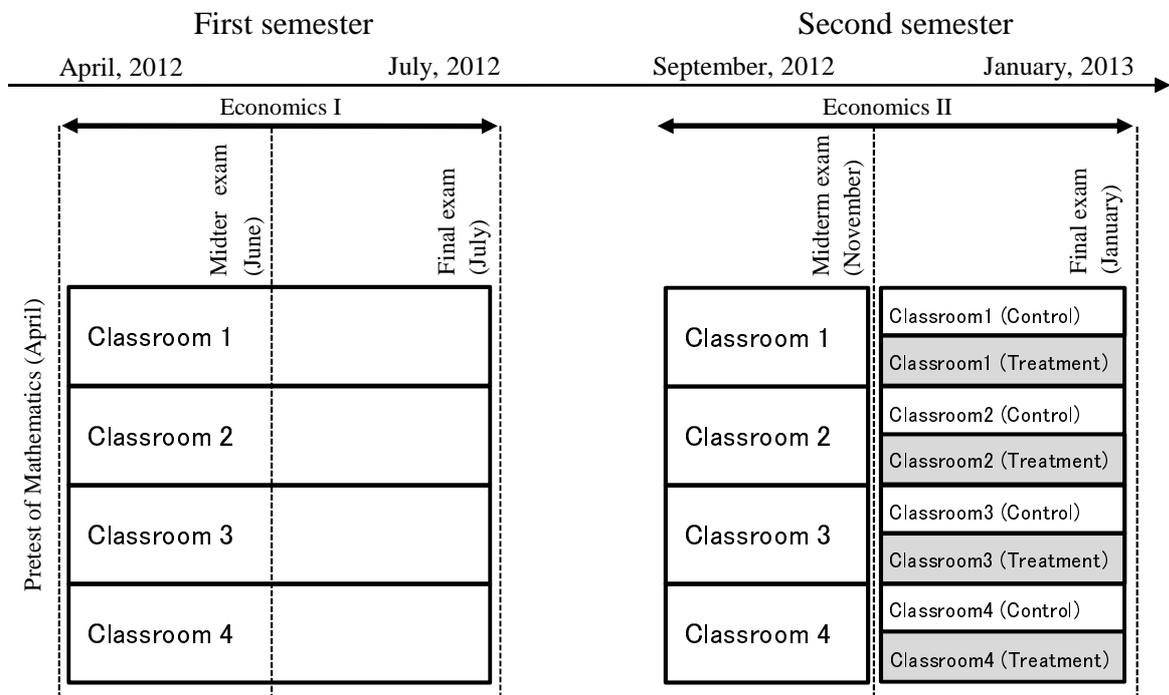
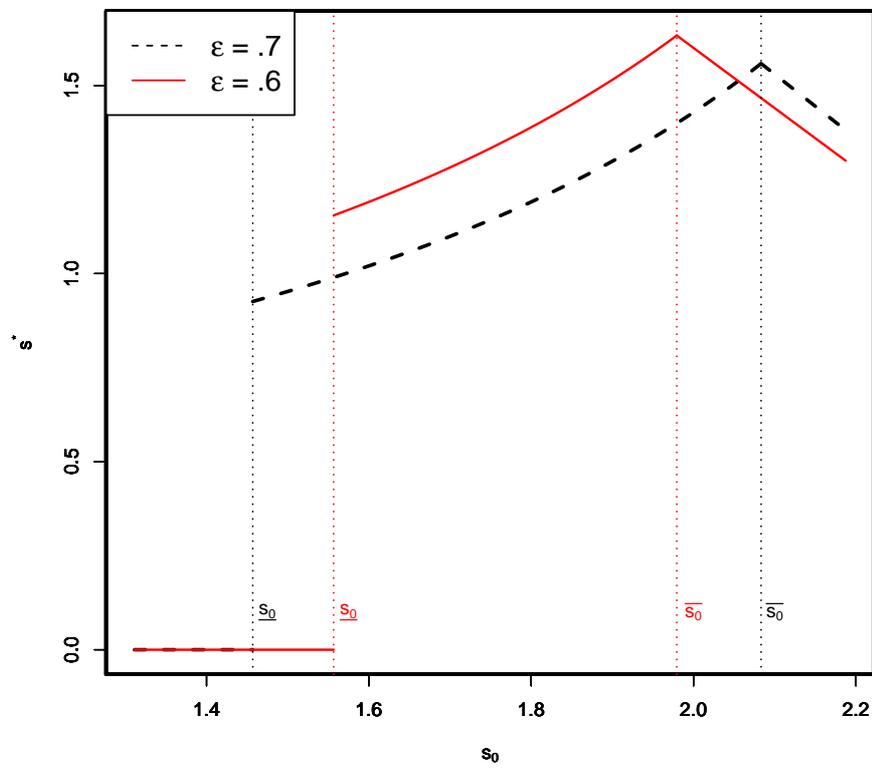
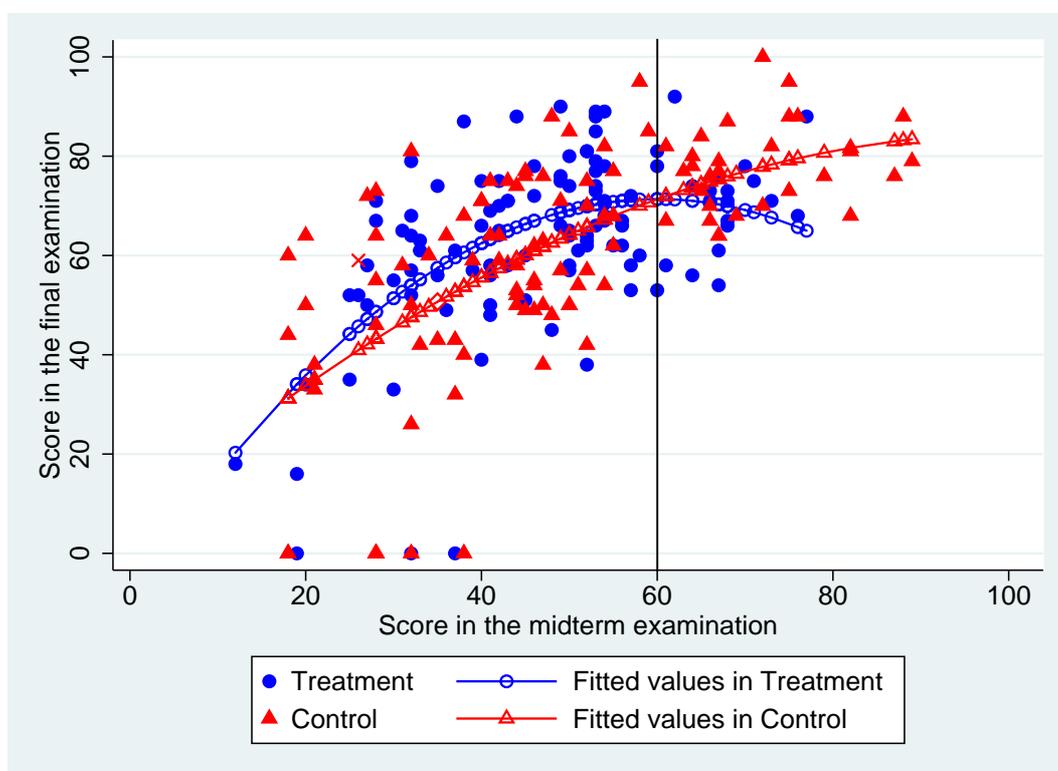


Figure 2: The relationship between midterm score (s_0) and equilibrium final score (s^*)



Note: We set parameter value $S = 3$.

Figure 3: The relationship between the midterm examination scores and improvements in the final examination scores



Note: We calculate the fitted values of final examination score using the coefficients in column (1) in Table 3. The student characteristics are evaluated at the sample mean values.

Figure 4: The letter to students (treatment group)

Introductory Economics, Second Semester, 2012
A report on the result of your midterm examination

Student ID	7 digit ID	Name	Name of the student
Classroom	Instructor's name		

Your score in the midterm examination is **49/110**
Of them, your score in problems of mathematics is **6/10**

Within four classrooms, you are **146th** out of 285 students.

Original (written in Japanese)

2012 年度経済学通論 2 中間テスト結果

学籍番号 12E1-XXX 氏名 XXXX
経済学通論 2 クラス (XXXX)

中間テストの点数 49 点 / 110 点満点
うち 数学出題分 6 点 / 10 点満点

学年全体でのあなたの順位 285 人中 146 位

Figure 5: The letter to students (control group)

Introductory Economics, Second Semester, 2012			
A report on the result of your midterm examination			
Student ID	7 digit ID	Name	Name of the student
Classroom	Instructor's name		
Your score in the midterm examination is			49/110
Of them, your score in problems of mathematics is			6/10

Original (written in Japanese)

2012 年度経済学通論 2 中間テスト結果	
学籍番号 12E1-XXX	氏名 XXXX
経済学通論 2 クラス (XXXX)	
中間テストの点数	49 点 / 110 点満点
うち 数学出題分	6 点 / 10 点満点