DISCUSSION PAPER SERIES

IZA

IZA DP No. 9565

Psychological Incentives, Financial Incentives, and Risk Attitudes in Tournaments: An Artefactual Field Experiment

Bram Cadsby
Jim Engle-Warnick
Tony Fang
Fei Song

December 2015

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Psychological Incentives, Financial Incentives, and Risk Attitudes in Tournaments: An Artefactual Field Experiment

Bram Cadsby
University of Guelph

Jim Engle-Warnick
McGill University

Tony Fang
Memorial University of Newfoundland, University of Toronto and IZA

Fei Song
Ryerson University

Discussion Paper No. 9565
December 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 9565
December 2015

ABSTRACT

Psychological Incentives, Financial Incentives, and Risk Attitudes in Tournaments: An Artefactual Field Experiment

Tournaments are widely used to assign bonuses and determine promotions because of the link between relative performance and rewards. However, performing relatively well (poorly) may also yield psychological benefits (pain). This may also stimulate effort. Through a real-effort artefactual field experiment with factory workers and university students as a comparison group in China, we examine how both psychological and financial incentives, together with attitudes toward risk, may influence motivation and performance. We provided performance-ranking information both privately and publicly, with and without rank-based financial incentives. Our results show that performance-ranking information had a significant motivational effect on average performance for students, but not for that of workers. Adding financial incentives based on rank provided little evidence of further improvement. Much of the difference between workers and students can be explained by differences in attitudes toward risk. Indeed, for both groups financial and psychological incentive effects are both inversely related to individual levels of risk aversion, and are positive and significant both for workers and for students who are sufficiently risk-tolerant.

JEL Classification: J30, J24, J33, C93, C91

Keywords: incentives, social comparison, performance feedback, peer pressure, tournament, risk aversion, artefactual field experiment

Corresponding author:

Tony Fang
Department of Economics
Memorial University of Newfoundland
St. John's, NL, A1C 5S7
Canada
E-mail: tfang@mun.ca

1. INTRODUCTION

Compensation is at the core of all employment exchanges. The relationship between employee compensation and both employee and firm performance has long been a focus of attention in management, economics, organizational studies, and sociology (e.g., Gomez-Mejia, and Welbourne, 1988; Jensen, 2003; Lazear, 1986, 2000; Zenger, 1992). The specific choice of compensation scheme has been linked to sorting and performance (Cadsby, Song, and Tapon, 2007; Gerhart and Milkovich, 1992; Gerhart and Rynes, 2003; Jensen, 2003; Lazear, 2000; Zenger, 1992). Recent decades have witnessed an increase in the prevalence of different forms of pay-for-performance in many organizations. Lazear and Shaw (2009) report that close to 67 percent of firms in the US use individual pay-for-performance incentives. According to WorldatWork's Private Company Incentive Pay Practices Survey (2012), the percentage of US for-profit companies with a pay-for-performance program rose from 43% in 2007 to 78% in 2012.

Pay-for-performance is usually tied to some measure of work output and/or firm performance (Lazear, 2000). The idea behind most pay-for-performance schemes is that pay levels should differ to reflect the different levels of contribution made by different individuals, thus aligning the incentives of employee agents with those of owner principals (Jensen and Murphy, 1990). However, many pay-for-performance systems face potential problems, including substantial costs associated with the measurement of performance and the availability and accuracy of such measures (Holmstrom and Milgrom, 1991; Nalebuff and Stiglitz, 1983). To overcome these problems, tournament-based compensation (TBC hereafter) is argued to be desirable when measures of relative performance are less costly and/or more reliable than absolute measures (Lazear and Rosen, 1981). By linking individual pay to relative performance, TBC also maintains control over the total wage bill, while motivating employees to exhibit greater effort and performance (Ehrenberg and Bognanno, 1990; Green and Stokey, 1983; Lazear, 1995; Pfeffer, 1994). In practice, TBC systems such as promotion of senior executives (O'Reilly, Main, and Crystal, 1988), perks and privileges rewarded to top managers (Eriksson, 1999), and prizes give to top athletes (Becker and Huselid, 1992; Ehrenberg and Bognanno, 1990) are widely available.

TBC systems may motivate people to perform well because of the link between relative performance and financial rewards. However, performing relatively well (poorly) may also yield psychological benefits (pain), which may also stimulate effort (Frank, 1985). Starting with Adam Smith, many have argued that psychological elements may be as important for understanding social and economic behavior as the strict rationality assumptions behind many economic theories (Smith, 1759). Recently, a great deal of evidence regarding social preferences based on field and laboratory research (e.g., Charness and Rabin, 1999 and the references therein) has provided the foundation for research on whether and to what degree workers respond to information regarding the productivity of their peers.

Mas and Moretti (2009) for example studied the performance of cashiers in a large US grocery chain. Among other things, they found that worker productivity responded positively to the arrival of a higher productivity worker who could observe the original worker's performance, especially when that worker's shift frequently overlapped with that of the higher productivity worker. They concluded that social pressure based on being observable to high productivity co-workers can increase performance. Similarly, Falk and Ichino (2006) used a controlled field experiment in Switzerland to demonstrate that workers stuffing envelopes for a fixed wage were more productive when working within sight of another worker in the same room than when working alone. In contrast, Bellemare, Lepage, and Shearer (2010), using a real-effort data-entry task, presented laboratory subjects in Canada with information about the actual output of a person who had previously performed the same task. They found no positive impact of such information on performance in either a fixed-wage or piece-rate setup. On the contrary, information that another person had done either very well or very poorly led to lower levels of output, particularly for male subjects under fixed-wage compensation.

Three recent studies examined the effect of non-financial recognition on performance.

Kosfeld and Neckermann (2011) showed that when subjects knew that top performers in a fixed-wage database project in Switzerland would receive a congratulatory letter, it significantly increased performance. Bradler, Dur, Neckermann and Non (2013) used a controlled field experiment involving data entry in Germany to show that unexpected recognition in the form of a thank-you card after two hours of work improves subsequent performance, particularly when only the best performers receive the recognition. Delfgaauw, Dur, Sol, and Verbeke (2009) ran a field experiment at a Dutch retail chain, employing a sales competition between the individual stores. The sales competitions were applied both with and without financial incentives. In both cases, the competition was associated with a significant growth in sales when the gender of the store manager was the same as that of a sufficiently large proportion of the store employees. There were no differences in performance improvement between the competitions with and without financial incentives.

A number of papers also recently examined the role played by the provision of ranking information under different compensation schemes. Under piece rates, Blanes i Vidal and Nossol (2011) showed that providing private ranking information can significantly improve performance in a firm setting in Germany when compared to a baseline in which such information was not provided. Azmat and Iriberri (2012) used a real-effort laboratory experiment in Spain to show that providing feedback on average group performance also resulted in improved individual performance under piece rates. Under fixed pay, such feedback produced an improvement in male performance but a deterioration in female performance with no significant effect overall. In contrast, Barankay (2012a) found that removing feedback on ranking improved performance of furniture salespeople paid by commission in a three-year field experiment in North America, while Barankay (2012b) found that piece-rate workers recruited through Mechanical Turk were less likely to return to work again after they had been ranked. In addition to Azmat and Iriberri (2012), Kuhnen and Tymula (2012) in the US and Charness, Masclet and Villeval (2014) in France examined the effect of relative performance feedback under fixed wages. Kuhnen and Tymula (2012) demonstrated that knowing one would receive private feedback on one's rank within a group of six to nine subjects either with certainty or with a 50% probability resulted in improved performance on a multiplication task compared to a control with no feedback. Charness, Masclet and Villeval (2014) showed that in three-person groups providing public feedback on the rank of all three group members in a decoding task improved performance. Remarkably, they also found that subjects were willing to spend money to improve artificially their own rank or to sabotage their competitors.

Relative performance feedback has also been studied in educational settings. Azmat and Iriberri (2010) studied the effect of providing class averages to Spanish high-school students, who had previously received solely their own grades. This resulted in a significant improvement in grades, which reverted to normal after the class averages were no longer provided. Lam, Yim, Law and Cheung (2004) examined the effect of announcing that private written information on one's performance rank would be provided to grade 7 students taking a typing course in Hong Kong upon completion of the course. While this resulted in better performance, it had a number of negative effects such as students sacrificing learning opportunities and choosing less challenging tasks. Tran and Zeckhauser (2012) used students studying English in Vietnam to examine the effect of providing both private and public ranking information on practice tests. Both treatments resulted in higher scores on the final standardized test compared to a control with no ranking information.

We conduct an artefactual field experiment with laboratory controls (Harrison and List, 2004) among factory workers in China to examine the effect of relative performance feedback within such a workplace. For comparison purposes, we also study the effect of such feedback using a laboratory experiment with undergraduate student subjects from the same region. In our experiment, subjects repeatedly perform a real-effort addition task for pay. In the control treatment, subjects receive a flat payment that was independent of performance. We employ a two-by-two factorial design to implement four additional experimental treatments: the first factor is private versus public feedback on one's relative rank in a group of 20 or 30; the second factor is whether pay is fixed and independent of rank or directly linked to it. Previous studies have found both private (e.g., Kuhnen and Tymula, 2012) and public (e.g., Charness, Masclet and Villeval, 2014) feedback on rank to be effective motivators relative to a control under fixed pay. To our knowledge, only Tran and Zeckhauser (2012) have undertaken a comparison of private versus public feedback. That comparison was done in the context of academic achievement. While public feedback appeared to be linked to higher academic achievement than private feedback, this result was only of marginal significance. Thus the authors highlight the need for further study of this issue. We do so here through a comparison of private versus public feedback under both fixed-pay and ranking-based compensation in a workplace setting, the first factor in our two-by-two design.

The second factor compares the effect of relative ranking feedback when pay is fixed and independent of ranking to the effect of pay being directly determined by one's relative ranking. This comparison isolates the potential psychological effects of ranking feedback on productivity from the combination of psychological and financial motivation if one's pay is directly linked to one's ranking. To our knowledge this has not been studied previously and it represents a major novel contribution of our paper.

During the experimental session, we measured subjects' aversion to uncertainty by measuring their risk preferences. Uncertainty can operate in a variety of ways in this setting: risk-averse subjects could simply avoid the lottery that requires high effort for an uncertain return and exert less effort for a more certain outcome. Moreover, aversion to uncertainty could manifest itself as discomfort, stress or an unpleasant and dysfunctional feeling about the imposed uncertainty of the ranking regime (Loewenstein, Hsee, Weber and Welch, 2001), which could adversely affect performance (Cadsby, Song and Tapon, 2007; 2009). Our hypothesis is thus that more risk-averse people are likely to improve less and in extreme cases even suffer a deterioration in performance, compared to less risk-averse people, who are likely to improve more under a ranking system. To our knowledge, the exploration of this issue is also a novel contribution of our study.

Our study brings together tournaments, peer pressure, field research and experimental control by introducing an artefactual field experiment, conducted in China with full-time factory workers, in which we compare and disentangle the effects of private versus public feedback on relative ranking, fixed versus graduated tournament pay, and attitudes toward risk on performance of a real-effort task. Chinese undergraduate students participate in an analogous laboratory experiment. The purpose of using both factory workers and university students is to explore whether and to what extent workers might behave and respond to incentives and risks differently than students, the typical participants in lab experiments, in the context of feedback on relative ranking. We find that they do behave differently, largely owing to differences in attitudes toward risk. We continue with the experimental design, behavioral predictions and experimental procedures. We then present the experimental results and draw some conclusions.

2. EXPERIMENTAL DESIGN

We used a standard real-effort task (e.g., Niederle and Vesterlunde, 2007), in which subjects computed the sum of a series of five two-digit numbers. Each number was drawn from the uniform distribution using a computerized random-number generator. The subjects solved these arithmetic problems over eight experimental rounds, with each round lasting three minutes. This particular task is ideal for the study because while both full-time employees and students can accomplish it, it may be taxing under the pressure of a time constraint. We conducted the experiment using paper-and-pencil so that computer proficiency was not an issue for any participant.

There were five experimental manipulations, which we denote conditions. The first condition, *fixed-pay*, is the control condition (hereafter denoted C), in which a participant earned ¥6.00 per round regardless of his/her performance and received no feedback on that performance. We use this control condition to compare output resulting from ability and effort exerted when performance has neither financial nor socio-psychological consequences with that resulting from ability and effort exerted under the different ranking/payment systems. For the four experimental conditions, we utilize a 2×2 factorial design that manipulates financial incentives (fixed pay vs. rank-based pay) and performance ranking information (privately vs. publicly revealed). Thus, the second condition is *non-financial with private ranking information* (NF:PR): each participant earned ¥6.00 per round regardless of his/her performance, and performance was ranked and reported privately to each individual participant. The third condition is *non-financial with public ranking information* (NF:PU*)*: each participant earned ¥6.00 per round regardless of his/her performance, and performance was ranked and announced publicly by reading out the names of the participants in the order of their performance rank. The fourth condition is *financial with private ranking information* (F:PR): each participant was paid based on his/her relative session performance ranking, and performance was ranked and reported privately. The fifth condition is *financial with public ranking information* (F:PU): each participant was paid based on his/her relative session performance ranking, and performance was ranked and announced publicly. Table 1 illustrates the tournament financial payoffs used in the 20-person worker sessions. A similar payoff structure was used for the 30-person student sessions.[1]

We conducted five experimental treatments, all of which consisted of eight rounds of play. In each of the five treatments, subjects experienced the control condition of fixed pay and no ranking information in both the first two and the last two rounds. Thus, the treatments differed only in the conditions the subjects faced during the middle four rounds. In the control treatment (C), participants were exposed to the fixed-pay with no ranking condition during the middle four rounds. In treatments NF:PR and NF:PU, participants received the non-financial with private ranking condition and non-financial with public ranking conditions respectively. In treatments F:PR and F:PU, participants received the financial private ranking and financial public-ranking conditions respectively. With this design, we are able to observe the change in within-person productivity from the first two rounds to the middle four rounds when the incentives are added, and then from the middle four rounds to the last two rounds when they are removed. We are then able to compare the observed changes in productivity between the various treatments. In addition, we are able to compare productivity levels between treatments by examining productivity in the

[1]The reason for the different session sizes is discussed in the Experimental Procedures section below.
[2]While this would seem more likely if the recognition were for one's everyday factory work rather than one's high ranking on an experimental real-effort task, recognition of a publicly announced high-ranking performance by one's peers and/or managers

middle four rounds across the five treatments.

3. BEHAVIORAL PREDICTIONS

3.1 Social Comparison

There are several strands in the social science literature that motivate our social comparison treatments. Frank (1985) notes local status, i.e., status within one's reference group, affects behavior. Frank (1985) argues that communicating one's ability to others is important in society, but since ability is unobservable it must be signaled through behavior. Moreover, Stajkovic and Luthans (2001) argue that social recognition, which is defined as personal attention to one's job performance by others, can affect behavior on the job. Such recognition functions as a non-financial reward based on job performance. Since it implies a positive evaluation of one's job performance by others, it may be predictive of future promotions or success. It can thus motivate better performance.[2] As discussed above, Mas and Moretti (2009) show that being observed by high-productivity colleagues increases performance through social pressure, and Falk and Ichino (2006) demonstrate that working independently but in pairs results in better performance than working alone. Furthermore, recent developments in neuroeconomics provide evidence that social comparison information has an immediate visible impact on motivational-related brain processes (Fliessbach et al., 2007; Dohmen et al., 2011).

Our NF:PU condition allows our subjects to signal their ability directly through their performance on the experimental task and receive the resulting social recognition. A comparison with the NF:PR condition permits us to measure the extent to which being able to signal one's status to others, a necessary condition to receiving their recognition, affects performance relative to the mere receipt of private social comparison information. The latter may also be motivating if the quality of one's performance is linked to one's level of self-satisfaction or ego utility even

[2]While this would seem more likely if the recognition were for one's everyday factory work rather than one's high ranking on an experimental real-effort task, recognition of a publicly announced high-ranking performance by one's peers and/or managers might nonetheless be regarded as a positive omen predictive of possible promotion or financial success.

when nobody else knows about it (e.g., Kuhnen and Tymula, 2012).[3] This leads to our first two behavioral hypotheses:

H1: Relative performance ranking information has motivational power on performance.

H2: Publicly announcing relative performance ranking will have greater motivational power on performance than privately conveying the same information to individuals.

Frank (1985) notes that concerns about one's position in the economic hierarchy can affect one's behavior. Clark, Friters, and Shields (2008) survey literature on relative income and happiness, and report empirical evidence of a positive relationship between them. Similarly, Knight, Song and Gunatilaka (2009) present empirical evidence from villages in rural China, demonstrating a significant relationship between income relative to one's fellow village residents and reported levels of happiness. Moreover, simply earning more even in absolute terms in the experiment clearly affords greater command over resources and hence higher utility according to standard economic theory. Our financial treatments allow us to test the additional effect of relative income linked to ranking, above and beyond mere ranking information with no financial consequences, on effort and performance. This leads to our third behavioral hypothesis:

H3: Financial benefits have an added motivational effect above and beyond the psychological benefits induced by relative-performance ranking information.[4]

3.2 Risk

Since it is generally impossible to predict one's relative performance with certainty, our experimental design leads to uncertainty for participants. A person who wishes to improve his/her rank faces psychological uncertainty as to whether an increase in effort will accomplish this, even in the absence of financial incentives. The effect is potentially magnified by the addition of financial uncertainty in the form of a financial incentive linked to rank. A subject who expends greater costly effort to improve his/her ranking is accepting a gamble involving a

[3]Besides demonstrating empirically that private ranking information can improve performance, Kuhnen and Tymula (2012) present an excellent discussion of the theoretical literature on ego utility.
[4]An alternative perspective is provided by the literature on incentive crowding-out effects (e.g., Frey and Jegen, 2001; Gneezy and Rustichini, 2000a; Gneezy and Rustichini, 2000b, Torgler, Frey, Schaffner and Schmidt, 2008).

certain increase in the disutility from effort to achieve a potentially better outcome.

Given that risk plays a central role in pay-for-performance compensation schemes, recent work has explored whether and to what extent one's attitude toward risk has an impact on the incentive property of PFP compensation schemes. The finding for piece-rate systems, in which a worker is financially rewarded in proportion to his/her output, is that the effectiveness of pay-for-performance compensation at improving productivity is inversely related to individual levels of risk aversion (Cadsby, Song and Tapon, 2007; 2009; Zubanov, 2012). In this paper we further explore this issue under circumstances in which workers are rewarded according to their *relative* performance, financially or non-financially. Specifically, we examine how individual attitudes toward risk may influence the size of the incentive effect of tournament-based ranking feedback where psychological and/or financial uncertainty exists.

H4: The effectiveness of tournament-based ranking incentives on improving performance is inversely related to individual levels of risk aversion.

4. EXPERIMENTAL PROCEDURES

4.1 Subject Pools

We held seventeen sessions with factory workers in Zhejiang, a province in southeast China, and six sessions with university students at Zhejiang Gongshang (Industry and Commerce) University in Hangzhou, capital of Zhejiang province. Table 2 summarizes the distribution of the sessions. As indicated in the table, we ran sessions with three different companies (denoted A, B, and C for privacy reasons). We ran at least one session of each treatment at each company, and we held two of each tournament pay session at Company A. Thus we have a total of three sessions each for treatments C, NF:PR and NF:PU and four sessions each for treatments F:PR and F:PU. There were 20 worker participants in each session for a total of 340 subjects. Worker participants were recruited with the assistance of managers in the case of two factories and union officials at another. We were assured that workers knew the other participants in their session well, and this certainly appeared to be the case. With the university student subject pool, we conducted two sessions each of treatments C, NF:PR, and NF:PU, and one session each of the financial incentive treatments. There were 40 students in the two control sessions combined and 30 in each of the experimental treatment sessions for a total of 220 student participants. The students in each of the experimental treatment sessions all came from one class, and thus knew each other well. Indeed the reason that we increased the size of the student sessions from the 20 participants used in the case of workers to 30 participants was to allow a whole class to be involved in each session. Social ranking and status is likely to matter more within a natural social group such as people working together in a small- to medium-sized factory or studying together in a Chinese university class.[5]

Table 3 provides a summary of the demographic characteristics of the participants. The mean age of the factory workers was about twenty-eight. Sixty-one percent of the subjects were female, and thirty-six percent were married. The mean monthly income was about ¥1700 (approximately $246.50 US), the mean education level was 2.66 (between junior and senior high school), and eighteen percent of the subjects were managers. The comparable characteristics for the student population was an average age of about twenty, fifty-six percent female, and an education level of five (university undergraduate). None were married.

4.2 Experimental Sessions

The experimental sessions were held in company conference rooms for employees and classrooms for students. Upon arrival, participants were seated apart from each other and no communication was permitted. The experimental instructions were read to the participants while they followed along on their own printed copies. Each participant was provided with a prepared workbook. For each round, the first page in the workbook explained which compensation scheme, or condition, would apply to the upcoming round. Figure 1 provides a sample workbook sheet translated into English for a representative experimental task. Participants were not permitted to look ahead to future pages or to go back to previous pages. They were allowed to

[5] In China, classes often take virtually all of their courses together for four years of university education. They also often live together in the same dormitories. This was the case at Zhejiang Gongshang University, the site of our study. Within the separate male and female dorms, four students from the same class shared one room. The rooms of other students from the same class were adjacent.

tear off one page and look at the next only when instructed to do so by the experimenter. After each round, each participant's workbook page was collected by the experimenters and taken to another room where the number of correct answers was calculated. In the four non-control treatments, according to the treatment they were in, participants received some or all of the following information in each of the middle four rounds before the next round started: feedback on their social comparative ranking information privately or publicly, and their earnings.

After participants completed the experimental task, they filled out a questionnaire in which they responded to a number of demographic questions such as age, gender, marital status and monthly income (the last two items only applied to workers). Besides collecting demographic data, another primary purpose of the questionnaire was to elicit risk preferences. We adopted a risk-aversion elicitation instrument based on Binswanger (1980) and Eckel and Grossman (2008). We find this instrument easy to understand, quick to administer, and thus appropriate to use when many of the worker participants had only junior high-school education or less.

Participants were shown ten circles, depicted in Figure 2. Each circle includes two payoffs, and each payoff occurs with a 50/50 chance. These ten circles are organized in a clock-wise fashion with the top circle containing two identical numbers, representing a certain payoff, while the subsequent eight circles represent lotteries that increase in both expected payoff and variance. The last two circles contain lotteries with identical expected payoffs. However, the last circle has a much higher variance to permit identification of participants who may not have risk-averse preferences. Participants were asked to indicate which one of the ten circles they would prefer to play. This risk measure instrument is advantageous for field use for at least two reasons. First, 50/50 gambles are easy to understand and expected payoffs are simple to calculate. Second, the measure is visually presented in a manner that focuses the attention of participants on the fact that the increase in expected earnings is associated with an increase in risk. This presentation has been used successfully among Peruvian farmers with limited education by Engle-Warnick, Escobal, and Laszlo (2009, 2011). After participants made and submitted their choices, the experimenter asked a volunteer participant to flip a coin in front of everyone to determine the payoff for each participant. The earnings of this task were added to the total session earnings for each participant. At the end of the session, players were taken individually to another room, where they were paid privately in cash.[6] Each session lasted just over an hour. On average, participants earned between ¥50 and ¥60 Yuan, which exceeded a day's pay for factory workers and was well in excess of the ¥10-¥15 an hour that students could earn at campus part-time jobs.

5. RESULTS

5.1 Aggregate Results

A preliminary look at the data revealed that results for a given treatment appeared to be quite similar at each of the three factory sites. To test for differences among the three sites we conducted a joint F-test on the following two regression models, where the dependant variable in the first regression is treatment productivity, TP (average number of problems solved in the middle four condition rounds) and the dependant variable for the second regression is treatment productivity improvement, TPI (the difference between the average number of problems solved in the middle four rounds and the number solved in the first two rounds):

$$TP,TPI = \beta_0 + \beta_1 \cdot D_1 + \beta_2 \cdot D_2 + \beta_3 \cdot T + \beta_4 \cdot T \cdot D_1 + \beta_5 \cdot T \cdot D_2. \qquad (1)$$

In the regressions, D_1 takes on the value 1 for site 2 and 0 otherwise, D_2 takes on the value 1 for site 3 and 0 otherwise, and T takes on the value 1 for the control treatment and 0 otherwise. The null hypothesis is H_0: $\beta_1 = \beta_2 = \beta_4 = \beta_5 = 0$. We ran eight regressions: both models for each of the four experimental treatments. We failed to reject the null of coefficient equality in any regression. Thus, we pooled the worker data for the subsequent analysis.

Table 4 presents average per-round productivity numbers for workers and students by treatment, where productivity is defined as the number of correctly completed summations. We focus on average per-round productivity in the middle four rounds in which the experimental manipulation is imposed and productivity improvement due to the manipulation, i.e the within-

[6] Experimental instructions and materials are available upon request.

person difference in average productivity between the middle four and first two rounds. We also report the difference in average productivity in the last two versus the middle four rounds. The final row of the table presents results from the lottery task. The responses were coded as integers, with the number 1 assigned to the safe choice at the top of the instrument, and each subsequent circle indexed by adding one, moving clockwise around the larger circle. Thus, a larger integer indicates a less risk-averse response.

The table reveals several broad stylized facts regarding the experimental results. First, even in the absence of performance-based incentives, productivity was significantly positive. Second, student productivity was always higher than worker productivity (broadly, students solved around 12 problems, while worker performance was in the neighborhood of seven). Third, across all experimental treatments both populations exhibited increases in productivity during the middle four rounds. However, while for the students this improvement was lowest in the control treatment as expected, for workers it was highest in that treatment. Fourth, workers made lottery decisions that were on average one full "circle" more risk-averse than students.

5.2 Treatment Effects

In Tables 5 and 6, we report statistical tests of our hypotheses. Table 5 focuses on treatment productivity in the middle four rounds, while Table 6 focuses on productivity improvement between the first two and the middle four rounds. In each case, panel A presents results for workers, while panel B presents results for students. Recall that treatment productivity is defined as the average number of problems correctly solved per round during the four middle rounds in which the experimental conditions are in effect, and productivity improvement is the difference between the average number of problems correctly solved per round in the middle four versus the first two rounds. In each case, there is one observation for each participant. The tests are performed in a regression framework with a random effect to control for unobserved differences between sessions.[7] In the column labeled Difference, we report the estimated average difference between the two treatments indicated in the column labeled Comparison Sets. We

[7]Analogous Wilcoxon rank-sum tests, not reported to save space, yield qualitatively identical results.

consider productivity improvement a better measure for incentive effects because it controls for individual differences in baseline ability and motivation that may affect productivity. However, we report and discuss both measures. We define three types of effects:

Incentive Effects: the difference between the control treatment C and each of the four experimental treatments in which various kinds of incentives are given;

Status Effects: the difference between the private and public ranking treatments: NF:PR − NF:PU and F:PR – F:PU;

Financial Incentive Effects: the difference between the non-financial and financial treatments: NF:PR − F:PR and NF:PU − F:PU.

In accordance with our unidirectional hypotheses H1 and H2, we employ one-tailed unidirectional tests to test for incentive and status effects. However, despite the unidirectional prediction made by H3, we employ two-tailed tests to examine the effects of financial incentive because of the possibility that the effects of psychological incentives are crowded out by the presence of financial incentives as mentioned in footnote 5.

Table 5, panel A indicates that there are no significant treatment effects of any kind for workers when average productivity in the middle four rounds is compared across treatments. Moreover Table 6, panel A also shows no significant treatment effects for workers using productivity improvement data. Although worker productivity improved in all four treatments, it improved by even more in the control treatment. Thus, the improvement was likely due to practice or learning. Moreover, none of our social comparison hypotheses appear to receive any support from the factory worker data.

Table 5, Panel B presents evidence for treatment effects on average productivity for the student population. First, t-tests show a statistically significant difference between the control treatment and all other treatments though this effect is of only marginal significance for the F:PU treatment (p = 0.09). Second, there is no evidence for status effects. Third, there is no evidence for financial incentive effects. Thus, only H1 received support, while H2 and H3 did not from the student average productivity data.

Table 6, Panel B presents evidence for treatment effects using within-subject productivity improvement for the students. In each case, this improvement is compared between two treatments involving different subjects, so it is also a between-subject comparison. In contrast with the factory employees, the student data showed that there was a significant incentive effect in three of the four experimental treatments when compared with the control. Moreover, public ranking information had a significant motivational impact in the financial but not in the non-financial treatments, while financial rewards had a marginal motivational impact ($p = 0.07$) in the public but not in the private treatments. Thus, H1, H2 and H3 received some support from the student productivity-improvement data.

On average, students seem to respond to ranking information, while workers do not. The upper portion of Table 7 reports a random-effects regression of Treatment Productivity Improvement on a dummy variable that is 1 for students and zero for workers for each of the five treatments. The coefficient on the student dummy is positive and significant for three of the four treatments in which ranking incentives were used during the middle four rounds, the exception being the F:PR treatment. Thus, while in the control treatment there was no significant difference between worker and student productivity improvement, in three of the incentive treatments student productivity improvement was significantly higher than worker improvement.

5.3 Risk Preference and Productivity Improvement

We now examine how subjects' response to performance-contingent incentives that come in both financial and non-financial forms may have been influenced by risk preference. Specifically, we hypothesize that the incentive effects of the four experimental treatments, measured by Treatment Productivity Improvement, may be correlated with individual levels of risk-aversion, and test this hypothesis by running the following regression:

$$\text{TPI} = \beta_0 + \beta_1 (\text{RA}) + \beta_2 (\text{Student Dummy}), \qquad (2)$$

where TPI is Treatment Productivity Improvement, RA is the level of risk-aversion, and the student dummy is 1 for students and 0 for factory employees. Again we use a random effect for sessions.

The results are summarized in the lower portion of Table 7.[8] They show that in every treatment, with the exception of the control, the RA coefficient is positive and significant. Since a larger RA index means a lower level of risk aversion, the interpretation is that TPI increases with decreasing risk aversion. This lends strong support to H4.[9] The student dummy remains positive and significant only for the F:PU treatment, suggesting that, controlling for risk attitude, students are more sensitive to the public revelation of ranking information than workers when ranking is linked directly to financial compensation. This is not surprising, as the students came from one class for each session, while the workers came from various parts of the factory. Thus, the students are likely to be a more coherent group than the factory workers. Hence, compared to the workers, they may feel more competitive with each other when they know their ranking and resulting earnings will be publicly announced. Notice however that for the other three treatments, the student dummies are not significant. Thus, unless ranking information is publicly announced and linked to financial rewards, differences between student and worker responses to such information disappear when risk preferences are taken into account.

5.4 Risk Preference and Treatment Interaction

Having established the correlation between risk preference and treatment productivity improvement, we next examine whether and to what extent risk-aversion may interact with the type of incentives present in a TBC setting to either induce or inhibit productivity improvement. We utilize the following regression model:

$$TPI = \beta_0 + \beta_1 (\text{Treatment}) + \beta_2 (\text{RA}) + \beta_3 (\text{Treatment}\cdot\text{RA}), \quad (3)$$

where TPI is Treatment Productivity Improvement, Treatment equals 0 for each of the four

[8]Gender was initially entered as a control variable. It was never significant and thus was dropped from the regressions.

[9]An alternative interpretation for the financial treatments might be that those who improved most felt wealthier and therefore when faced with the choice of a gamble at the end of the game were more likely to choose a risky gamble. While this endogeneity problem could theoretically arise in the two financial treatments, it should not be an issue in the non-financial treatments where payment does not depend on performance. All four treatments demonstrate a strong relationship between risk attitude and productivity improvement when exposed to the incentives in the middle rounds. The risk-aversion coefficients in the financial treatments (0.19 and 0.22) are almost identical to those in the non-financial treatments (0.21 and 0.20), and statistically there was no significant difference between them. Moreover according to a Kolmogorov-Smirnov test, there are no significant differences in the distribution of risk attitudes between the financial and non-financial treatments. All of these results are available upon request.

experimental treatments and 1 for the control treatment, and Treatment·RA is an interaction term. We run this regression separately for each of the four experimental treatments, again employing a random effect for sessions. In each case, we compare productivity improvement from the first two to the middle four rounds between one of the incentivized treatments and the control.

Before we proceed to this set of analyses, it is important to note that the median levels of risk-aversion for workers and students were quite far apart (5 and 8 respectively), while the modal choice was 9 for both workers and students. Table 8 presents a more detailed picture of the risk-aversion data we gathered from both workers and students. We centered the risk-aversion index at each of the 10 possible risk-aversion levels for regressions using Model (3) in order to provide an explicit picture of the marginal effect of the treatment relative to the control at each risk-aversion level.[10]

Results of this analysis are summarized in Table 9 for workers and Table 10 for students. Both tables are organized horizontally by experimental treatment. Within each treatment we present estimated coefficients for Model (3), with the risk-aversion index centered at each possible value in turn. Centering the risk-aversion index at different values does not affect the estimated coefficients on either the risk-aversion level or the interaction between risk-aversion level and treatment. In each of the treatments for workers and students, the estimated coefficient on risk-aversion level is positive. It is significant in all cases except for the F:PR treatment for students where the coefficient is of the predicted sign but not significant. This indicates that in each of the incentivized treatments with the exception of F:PR for students, productivity improvement is significantly higher for people who are less risk-averse and thus chose a gamble indexed to a higher number. In each of the treatments for workers and students including F:PR, the estimated coefficients on the interaction term were negative and significant. This means that the effect of risk aversion on productivity improvement was significantly lower in the control than in the experimental incentivized treatments.

[10] Cohen, Cohen, West and Aiken (2003: 261–282, especially p. 281) contain an excellent discussion about centering and the proper interpretation of centered versus non-centered variables.

Centering the risk-aversion index at different levels is useful because of its effect on the estimated constant term and the estimated coefficient on the treatment dummy. The estimated constant term can be interpreted as the marginal improvement in productivity in the incentivized treatment for participants whose risk preferences correspond to the centered number. For example, for treatment NF:PR, with risk aversion centered at 8, the productivity improvement is predicted to be 1.49 (significant at the 1% level) for workers whose risk preferences were at that level. The coefficient on the treatment effect indicates the difference between the increase in productivity in the control versus the incentivized treatment, the focus of our analysis. For example, for treatment NF:PR, with risk aversion centered at 8, the coefficient estimate of the treatment effect is −0.63 (significant at the 5% level) for workers. The interpretation is that for workers whose selected risk-aversion level was 8, the productivity improvement was significantly lower under the control than under non-financial, private ranking treatment. Notice that this is true despite the fact that, as previously discussed, on average there was no such significant effect for workers when risk preferences were ignored.

The implications of this analysis are important. If we ignore risk preferences, and examine only the average response, students seem to respond significantly to ranking information, while workers do not. However, once we control for risk preferences, we find that the privately announced rankings resulted in higher productivity improvement for workers than in the control treatment when their aversion to risk was sufficiently small. In the case of publicly announced rankings with no financial implications, the effect is there for the least risk-averse workers only, and the significance is only at the 10% level. However, recall that 23% of workers were in this category. Paying workers according to publicly announced rankings did not yield significant treatment effects for any workers compared to the control. For students, all of the incentivized treatments were significant relative to the control for those who were sufficiently risk-averse.[11, 12]

[11]We also examined whether subjects were reacting to the announcement that a new information/compensation environment would apply made at the beginning of period 3 as assumed throughout this section and/or to the receipt of ranking information (and consequent earnings in the financial treatments) that first occurs at the end of period 3. We could find no systematic reaction

5.5 Removing the Incentives

In the last two rounds of each treatment, all psychological and financial incentives were removed. As in the first two rounds, the subjects were now paid a fixed amount that did not depend on performance. In addition, ranking information was no longer provided. We did not develop a formal hypothesis about what would happen when the incentives were removed. If the incentives primarily affected the subjects through motivating them simply to work harder or at a faster pace, one might expect that subjects would revert to lower levels of effort once the incentives were withdrawn. However, if the incentives caused subjects to focus on learning how to perform the task more effectively, the effect of the incentives might outlive their presence. The latter was indeed the case. In no case was there a significant difference between average per-round productivity in the middle four rounds and average per-round productivity in the last two rounds for workers or for students.[13]

6. GENERAL DISCUSSION

We undertook an artefactual field experiment in which we examined the effect of feedback on performance ranking both when such ranking was linked to financial compensation and when it was not. Our experiments enable us to examine whether and to what extent two aspects of rank-based incentives, financial and psychological, independently influence motivation and performance; whether the effect of publicly announcing worker rankings differs from the effect of privately revealing individual ranking information; how attitude toward risk influences the magnitude of ranking incentive effects for individuals; and whether students and full-time workers react to ranking incentives differently. We found that subjects produce positive amounts even in the complete absence of performance-based incentives. We also discovered robust evidence for the effects of the ranking information treatments among students. On average,

---

to the receipt of ranking information or payments based on it. The observed productivity effects all occur from period 3 onward based on the announcement that a new information/compensation condition would henceforth be taking effect.

[12]We also examined whether individual risk attitudes affected the status and financial incentive effects reported in Table 6. They did not. The results are not reported to save space, but are available from the authors upon request.

[13]The detailed results are not reported here in order to save space, but are available from the authors upon request. Moreover, there was no significant relationship between change in productivity when the incentives were removed and attitudes toward risk. Again the results are available upon request.

the student participants performed significantly better during the middle four rounds of the experimental treatments, which were the incentivized rounds, than in those of the control. Moreover, student improvement in productivity from the initial two non-incentivized rounds to the middle four rounds with incentives was significantly greater than in the control in both non-financial as well as in the financial treatment with public ranking announcements. These treatment effects persisted even after the treatments were removed. However, on average, the factory workers did not show such a response. Thus, students, the usual subjects of laboratory experiments, appeared to respond positively to ranking incentives, while factory workers did not.

We found that the incentive effects on productivity improvement were significantly stronger when ranking information was made public for students in the financial treatments. There were no such effects for factory workers. However, adding financial incentives to the psychological incentives from receiving ranking information resulted in greater productivity improvement only for students when rankings were publicly announced. This result was only marginally significant. This suggests the psychological incentives alone were sufficient to motivate students in the private ranking treatments and workers in both the private and public ranking treatments to perform at their best, rendering performance-based financial incentives superfluous in these cases.

The observed differences between worker and student responses to incentives were partially reconciled by examining individual attitudes toward risk and uncertainty, and how such attitudes relate to individual responses to ranking incentives. Workers were on average significantly more risk-averse than students. This is an empirical observation, which we did not predict. We can only speculate that it may reflect the lower education level (between junior and senior high school on average with only 5.4% possessing at least some undergraduate education) and consequent lower expected lifetime earnings of many of the factory workers in our sample, when compared to the students, who were on their way toward the completion of an undergraduate degree. If risk aversion falls on average with higher wealth or expected lifetime earnings, this could potentially account for the differing distribution of risk attitudes between the two populations. However, we could not find any relationship between risk attitudes of individual workers and their current incomes, marital status, ages, or reported education levels.

6.1 Theoretical and Practical Contributions

Our paper extends the literature in several important directions. Theoretically, our results suggest that response to financial incentives depends not only on context, but also on individual heterogeneity, and in particular on individual attitudes toward risk. Thus, our paper extends and contributes to the literature on the relationship between risk attitudes and one's behavioral response to incentives. For example, it is consistent with earlier studies using piece-rate pay-for-performance systems and various real-effort tasks with student samples (Cadsby, Song and Tapon, 2007; 2009). More risk-averse individuals, whether students or workers, were significantly less responsive to both financial and psychological incentives relative to the control than those who were less risk-averse. Furthermore, less risk-averse participants, whether workers or students, performed significantly better in three of the four experimental treatments than in the control, and students performed better in the fourth one as well. In both private ranking treatments and the non-financial public-ranking treatment, there were no significant differences in performance improvement between students and workers after controlling for individual levels of risk-aversion. However, in the public ranking treatment with financial incentives linked to rank, individual levels of risk aversion only partially accounted for the different responses.

Prior research on the effects of tournament incentive and feedback has mainly used student samples (e.g., Azmat and Iriberri, 2012; Charness, et al., 2014; Kuhnen and Tymula, 2012). Conducting an artefactual field experiment using factory workers along with a laboratory experiment using the standard student subjects enabled us to gain considerable insight into how and why these two populations might differ in their responses to ranking information, whether or not such information is linked to pay. The differing distributions of attitudes toward risk between the worker and student populations are important because the magnitude of the incentive effects was inversely related to individual levels of risk aversion for both workers and students. Such differences demonstrate the need to move beyond the use of students as subjects in laboratory experiments involving work issues. Comparing our two subject pools reveals that performance feedback on relative ranking can be a useful tool to encourage high performance, but the effectiveness of such a program depends critically on the characteristics of the work force.

Our paper offers important managerial implications as well. Together with Cadsby et al. (2007), we demonstrate that individual attitudes toward risk/uncertainty can significantly impact how effective an incentive scheme may be in improving productivity. In this paper, we show that such an impact occurs not only with the introduction of financial incentives, but also with the use of feedback information on one's ranking whether or not such information is related to pay. In particular, when given feedback on one's performance ranking, more risk-averse employees are likely to be less responsive, unresponsive, or even respond in the wrong direction. Since the stakes are bound to be higher in the workplace than in our experiment, and risk aversion is likely to be greater at higher stake levels (Holt and Laury, 2002), this phenomenon could well be even more pronounced in the workplace than in our study. Thus, different kinds of pay schemes may suit different kinds of workers, and risk attitudes may be a critical factor in determining the best employee-compensation fit.

6. 2 Limitations and Directions for Future Research

Several features of our empirical investigative strategy deserve attention. First, the experimental set-up necessarily requires the elimination or exertion of control over all possible influences on performance other than the two key features on which we focus: private versus public provision of feedback on one's performance ranking and whether or not one's performance ranking is linked to one's pay. The resulting experimental environment and the somewhat abstract nature of the experiment may limit the direct generalizability of the findings to more complex organizational and economic settings where tournaments are used to motivate effort and productivity. Thus, although the experimental design maximizes the chance of detecting causally valid relationships between incentives based on ranking information and task performance, it risks reduced external validity. Notwithstanding these limitations, the methodology employed in our research has the strong advantage of creating a controlled laboratory environment on the factory floor, enabling us to abstract from confounding factors in order to isolate and disentangle the financial and psychological incentives of feedback information on productivity and explore its underlying moderating mechanisms via one's attitudes toward risk and uncertainty.

Future research should continue this line of inquiry and test robustness of the inverse relationship between risk-aversion and incentivized productivity improvement. For example, a long-term field study would be beneficial to examine whether the effects observed in this study are short- or long-term in nature. A second issue worthy of study is to examine the underlying mechanisms that lead to the inverse relationship between aversion to risk and responsiveness to financial and non-financial incentives. For example, it is important to explore further whether the themes examined in this paper are related to the psychological phenomenon of "choking-under-pressure" (e.g. Baumeister, 1984; Baumeister and Showers, 1986). Attitudes toward risk and uncertainty do not appear explicitly in the "choking-under-pressure" literature. However, our empirical observations suggest that there may be a relationship with those who dislike risk/uncertainty more likely to choke when forced to perform in a risky or uncertain environment.

References:

Azmat, G. and Iriberri. N. 2010. The importance of relative performance feedback information: Evidence from a natural Experiment using high school students. *Journal of Public Economics*, 94, 435–452.

Azmat, G. and Iriberri. N. 2012. The provision of relative performance feedback information: An experimental analysis of performance and happiness. CEP Discussion Paper 116, Centre for Economic Performance, LSE.

Barankay, I., 2012a. Rank incentives: Evidence from a randomized workplace experiment. Working Paper, Wharton School of Business, University of Pennsylvania.

Barankay, I., 2012b. Ranking and social tournaments: Evidence from a crowd-sourcing experiment. Working Paper, Wharton School of Business, University of Pennsylvania.

Baumeister, R. F. 1984. Choking under Pressure: Self-Consciousness and Paradoxical Effects of Incentives on Skillful Performance. *Journal of Personality and Social Psychology*, 46(3), 610–620.

Baumeister, R. F., and Showers, C. J. 1986. A Review of Paradoxical Performance Effects: Choking under Pressure in Sports and Mental Tests. *European Journal of Social Psychology*, 16(4), 361–383.

Bellemare, C., Lepage, P. and Shearer, B. 2010. Peer pressure, incentive effects, and gender: An

experimental analysis of motivation in the workplace. *Labour Economics*, 17, 276–283.
Becker, B. E., and Huselid, M. A.1992. The incentive effects of tournament compensation systems. *Administrative Science Quarterly*, 37, 336–350.
Binswanger, H. P. 1980. Attitudes toward risk: Experimental measurement in rural India. American. *Journal of Agricultural Economics*, 62, 395–407.
Blanes i Vidal, J. and Nossol, M. 2011. Tournament without prizes: Evidence from personnel records. *Management Science*, 51, 1721–1736.
Bradler, C., Dur, R. and Neckermann, S. and Non, A. 2013. Employee recognition and performance: A field experiment. *CESifo Working Paper Series 4164*, CESifo Group Munich.
Cadsby, C. B., Song, F. and Tapon, F. 2007. Sorting and incentive effects of pay for performance: An experimental investigation. *Academy of Management Journal*, 50, 387–405.
Cadsby, C. B., Song, F. and Tapon, F. 2009. The Impact of Risk Aversion and Stress on the Incentive Effect of Performance Pay. Working Paper 2009-12, University of Guelph.
Charness, G. and Rabin, M. 2002. Understanding social preferences with some simple tests. *Quarterly Journal of Economics*, 117, 817–869.
Charness, G., Masclet, D. and Villeval, M. C. 2014. The dark side of competition for status. *Management Science*, 60, 38–55.
Clark, A., Frijters, P. and Shields, M. 2008. Relative income, happiness, and utility: An explanation for the Easterlin paradox. *Journal of Economic Literature*, 46, 95–144.
Cohen, J., Cohen, P., West, S. and Aiken, L. 2003. *Applied Multiple Regression /Correlation Analysis for the Behavioral Sciences*, 3rd edition. Mahwah, NJ: Lawrence Erlbaum Associates.
Delfgaauw, J., Dur, R., Sol, J. and Verbeke, W. 2013. Tournament incentives in the field: Gender differences in the workplace. *Journal of Labor Economics*, 31, 305–326.
Dohmen, T., Falk, A., Fliessbach, K., Sunde, U. and Weber, B. 2011. Relative versus absolute income, joy of winning, and gender: Brain imaging evidence. *Journal of Public Economics*, 95, 279–285.
Eckel, C. C. and Grossman, P. J. 2008. Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior and Organization*, 68, 1–17.
Ehrenberg, R. G. and Bognanno, M. L. 1990. Do Tournaments Have Incentive Effects? *Journal of Political Economy*, 98, 1307–1324.
Engle-Warnick, J., Escobal, J. and Laszlo, S. 2009. How do additional alternatives affect individual choice under uncertainty? *Canadian Journal of Economics*, 42, 113–140.
Engle-Warnick, J., Escobal, J. and Laszlo, S. 2011. Ambiguity aversion and portfolio choice in small-scale Peruvian farming. *B.E. Journal of Economic Analysis and Policy*, 11 (1): Article 68.
Eriksson, T., Poulsen, A. and Villeval, M. 2009. Feedback and incentives: Experimental evidence. *Labour Economics*, 16, 679–688.
Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C., and Falk, A. 2007. Social comparison affects reward-related brain activity in the human ventral striatum. *Science*, 318, 1305–1308.
Falk, A. and Ichino, A. 2006. Clean evidence on peer pressure. *Journal of Labor Economics*, 24, 39–57.
Frank, R. 1985. *Choosing the Right Pond*. New York: Oxford University Press.
Frey, B. and Jegen, R. 2001. Motivation Crowding Theory. *Journal of Economic Surveys*, 15, 589–611.

Gerhart, B., and Milkovich, G. T. 1992. Employee compensation: Research and practice. In M. D. Dunnette, & L. M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (2nd ed.), Vol. 3: 481– 570, Palo Alto, CA: Consult. Psychol. Press.
Gerhart, B. and Rynes, S. 2003. *Compensation: Theory, Evidence, and Strategic Implications*. Thousand Oaks, CA: Sage.
Gomez-Mejia, L. R. and Welbourne, T. M. 1988. Compensation strategy: An overview and future steps. *Human Resource Planning*, 11, 173–190.
Gneezy, U. and Rustichini, A. 2001a. Pay enough or don't pay at all. *Quarterly Journal of Economics*, 115, 791–810.
Gneezy, U. and Rustichini, A. 2001b. A fine is a price. *Journal of Legal Studies*, 29, 1–18.
Green, J. R. and Stokey, N. L. 1983. A comparison of tournaments and contracts. *Journal of Political Economy*, 91, 349–364.
Harrison, G. W. and List, J. A. 2004. Field experiments. *Journal of Economic Literature*, 42, 1009–55.
Holmstrom, B. and Milgrom, P. 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics and Organization*, 7, 24–52, Special I.
Jensen, M. C. 2003. Paying people to lie: The truth about the budgeting process. *European Financial Management*, 9, 379–406.
Jensen, M. C. and Murphy, K. 2003. Performance pay and top-management incentives. *Journal of Political Economy*, 98, 225–264.
Knight, Song and Gunatilaka, 2009. Subjective well-being and its determinants in rural China. *China Economic Review*, 20, 635–649.
Kosfeld, M. and Neckermann, S. 2011. Getting more work for nothing? Symbolic awards and worker performance. *American Economic Journal: Microeconomics*, 3, 86–99.
Kuhnen, C. and Tymula, A. 2012. Feedback, self-esteem, and performance in organizations. *Management Science*, 58, 94–113.
Lam, S., Yim, P., Law, J. and Cheung, R. 2004. The effects of competition on achievement motivation in Chinese classrooms. *British Journal of Educational Psychology*, 74, 281–296.
Lazear, E. P. 1986. Salaries and piece rates. *Journal of Business*, 59, 405–432.
Lazear, E. P. 2000. The power of incentives. *American Economic Review*, 91, 410–415.
Lazear, E. P. and S. Rosen. 1981. "Rank order tournaments as optimum labor contracts. *Journal of Political Economy*, 89, 841–64.
Lazear, E. and Shaw, K. 2009. Wage structure, raises and mobility: An introduction to international comparisons of the structure of wages within and across firms. In Edward Lazear and Kathryn Shaw (eds.), *The 20 Structure of Wages: An International Comparison*. University of Chicago Press.
Loewesntein, J., Weber, E., Hsee, C. and Welch, N. 2001. Risk as feelings. *Psychological Bulletin,* 127, 267–286.
Mas, A. and Moretti, E. 2009. Peers at work. *American Economic Review,* 99, 112–45.
Nalebuff, B. and Stiglitz, J. E. 1983. Prizes and incentives: Towards a general theory of compensation and competition. *Bell Journal of Economics*, 14, 21–43.
Niederle, M. and Vesterlund, L. 2007. Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 122, 1067–1101.
O'Reilly, C., Main, B. and Crystal, G. 1988. CEO Compensation as Tournament and Social Comparison: A Tale of Two Theories. *Administrative Science Quarterly,* 33, 257–274.

Pfeffer, J. 1994. *Competitive Advantage through People: Unleashing the Power of the Work force*, Boston, MA: Harvard Business School Press.
Prendergast, C. 1999. The provision of incentives in firms. *Journal of Economic Literature*, 37, 7–63.
Smith, A. 1759 [1976]. *The Theory of Moral Sentiments*. London: Oxford University Press. First edition.
Stajkovic, A. D. and Luthans, F. 2001. Differential effects of incentive motivators on work performance. *Academy of Management Journal*, 44, 580–590.
Torgler, B., Frei, B., Schaffner, M. and Schmidt, S. 2008. A crowding-out effect for relative income. CREMA Working Paper 2008–20, Center for Research in Economics, Management and the Arts.
Tran, A., and Zeckhauser, R. 2012. Rank as an inherent incentive: Evidence from a field experiment. *Journal of Public Economics*, 96, 645–650.
Vandegrift, D. and Yavas, A. 2009. Men, women, and competition: An experimental test of behavior. *Journal of Economic Behavior and Organization*, 72, 554–570.
WorldatWork 2012. Private Company Incentive Pay Practices Survey. A report by WorldatWork and Vivient Consulting. Available online: http://www.worldatwork.org/waw/adimLink?id=58146
Zenger, T. R. 1992. Why do employees only reward extreme performance? Examining the relationship among performance pay and turnover. *Administrative Science Quarterly***,** 37, 198–219.
Zubanov, N. 2012. Risk aversion and effort in an incentive pay scheme with multiplicative noise: Theory and experimental evidence. Working paper. Erasmus University.

Table 1 Payoff table according to the relative ranking in a 20-person session

Ranking	Earnings
1	¥11.70
2	¥11.10
3	¥10.50
4	¥9.90
5	¥9.30
6	¥8.70
7	¥8.10
8	¥7.50
9	¥6.90
10	¥6.30
11	¥5.70
12	¥5.10
13	¥4.50
14	¥3.90
15	¥3.30
16	¥2.70
17	¥2.10
18	¥1.50
19	¥0.90
20	¥0.30

Table 2 Summary of sessions conducted at different locations

Treatment / Research Site	C	NF:PR	NF:PU	F:PR	F:PU
Company A	1	1	1	2	2
Company B	1	1	1	1	1
Company C	1	1	1	1	1
University	2	2	2	1	1

Table 3 Participants' demographic characteristics

Variable	Workers (n=340)		Students (n=220)	
	Mean	Std. Dev.	Mean	Std. Dev.
Age	28.30	10.05	19.66	0.76
Female	0.61	0.49	0.56	0.50
Married	0.36	0.48	0.00	0.00
Monthly Income	1697.57	801.05	N/A	N/A
Education Level*	2.66	0.94	5.00	0.00
Manager	0.18	0.36	N/A	N/A

Note: Education Level coding: elementary school =1, junior high=2, senior high=3, college=4, university=5, post-graduate=6.

Table 4: Data overview

Variable		Workers (n=340)		Students (n=220)	
		Mean	Std. Dev.	Mean	Std. Dev.
Average per-round productivity in the middle four rounds	C	7.05	2.89	10.31	4.44
	NF:PR	6.65	3.19	12.66	3.62
	NF:PU	7.02	3.26	12.02	4.04
	F:PR	7.08	2.70	12.81	3.67
	F:PU	7.33	3.37	11.96	3.41
Average per-round productivity improvement (Within-person Difference between middle four and first two rounds)	C	0.63	1.25	0.61	1.42
	NF:PR	0.84	1.17	1.50	1.86
	NF:PU	0.67	1.18	1.31	1.46
	F:PR	0.64	1.35	0.99	1.40
	F:PU	0.71	1.34	1.98	1.93
Difference between average per-round productivity in the last two rounds and the middle four rounds	C	−0.30	1.31	−0.18	1.55
	NF:PR	0.43	1.14	0.66	2.19
	NF:PU	0.59	1.06	0.35	1.35
	F:PR	0.23	1.43	0.44	2.10
	F:PU	0.19	1.49	−0.18	3.13
Risk Aversion		5.13	3.25	6.12	3.07

Table 5: Between-subject treatment effects based on average productivity in the middle four rounds

Panel A: Results for Workers

	Comparison Sets	Coef.	S.E.	Comparison Sets	Coef.	S.E.
Incentive Effects§	C − NF:PR (n=120)	0.40	1.09	C − NF:PU (n=120)	0.38	1.54
	C − F:PR (n=140)	−0.03	0.92	C − F:PU (n=140)	−0.27	1.17
Status Effects§	NF:PR − NF:PU (n=120)	-0.36	1.33	F:PR − F:PU (n=160)	-0.24	0.83
Financial Incentive Effects	NF:PR − F:PR (n=120)	-0.39	1.20	NF:PU − F:PU (n=140)	−1.27	1.15

Panel B: Results for Students

	Comparison Sets	Coef.	S.E.	Comparison Sets	Coef.	S.E.
Incentive Effects§	C − NF:PR (n=100)	−2.35***	0.80	C − NF:PU (n=100)	−1.70***	0.85
	C − F:PR (n=70)	−2.50***	0.99	C − F:PU (n=70)	−1.65*	0.97
Status Effects§	NF:PR − NF:PU (n=120)	0.75	0.70	F:PR − F:PU (n=60)	0.85	0.91
Financial Incentive Effects	NF:PR − F:PR (n=90)	0.15	0.81	NF:PU − F:PU (n=90)	0.06	0.86

A separate regression is run for each comparison with average productivity in the middle four rounds as the dependent variable. Each regression compares two treatments in a model with a constant term and a treatment dummy variable equal to 1 for the treatment to the left of the minus sign and zero for the treatment to the right of the minus sign. A random-effect specification is used to control for unobserved session effects. The coefficient on the constant term is not reported to save space. § denotes one-tailed tests as explained in the text. ***, **, and * denote significance at 1%, 5% and 10% respectively.

Table 6: Between-subject treatment effects using within-person productivity improvement data from the first two to the middle four rounds

Panel A: Results for Workers

	Comparison Sets	Coef.	S.E.	Comparison Sets	Coef.	S.E.
Incentive Effects§	C − NF:PR (n=120)	−0.21	0.26	C − NF:PU (n=120)	−0.03	0.26
	C − F:PR (n=140)	−0.01	0.22	C − F:PU (n=140)	−0.08	0.22
Status Effects§	NF:PR − NF:PU (n=120)	0.17	0.30	F:PR − F:PU (n=160)	−0.07	0.21
Financial Incentive Effects	NF:PR − F:PR (n=120)	0.18	0.26	NF:PU − F:PU (n=140)	−0.07	0.23

Panel B: Results for Students

	Comparison Sets	Coef.	S.E.	Comparison Sets	Coef.	S.E.
Incentive Effects§	C − NF:PR (n=100)	−0.90***	0.34	C − NF:PU (n=70)	−0.70**	0.29
	C − F:PR (n=100)	−0.39	0.34	C − F:PU (n=70)	−1.37***	0.40
Status Effects§	NF:PR − NF:PU (n=120)	0.20	0.31	F:PR − F:PU (n=60)	−0.98**	0.43
Financial Incentive Effects	NF:PR − F:PR (n=90)	0.51	0.38	NF:PU − F:PU (n=90)	−0.67*	0.36

A separate regression is run for each comparison with within-person productivity improvement from the first two to the middle four rounds as the dependent variable. Each regression compares two treatments in a model with a constant term and a treatment dummy variable equal to 1 for the treatment to the left of the minus sign and zero for the treatment to the right of the minus sign. A random-effect specification is used to control for unobserved session effects. The coefficient on the constant term is not reported to save space. § denotes one-tailed tests as explained in the text. ***, **, and * denote significance at 1%, 5% and 10% respectively.

Table 7: Population and risk attitude as determinants of productivity improvement

Treatment		C (n=100)	NF:PR (n=120)	NF:PU (n=120)	F:PR (n=110)	F:PU (n=110)
Model 1	Constant	0.63***	0.84***	0.67***	0.64***	0.71***
		(0.27)	(0.28)	(0.29)	(0.15)	(0.17)
	Student	−0.03	0.67**	0.64**	0.35	1.26***
	Dummy	(0.17)	(0.20)	(0.20)	(0.29)	(0.33)
Model 2	Constant	0.64**	−0.17	−0.30	−0.32	−0.56*
		(0.016)	(0.30)	(0.25)	(0.23)	(0.30)
	Risk	−0.002	0.21***	0.20**	0.19***	0.22***
	Aversion	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)
	Student	−0.02	0.23	0.44	0.28	1.17***
	Dummy	(0.27)	(0.33)	(0.28)	(0.29)	(0.29)

A separate regression is run for each treatment. The dependent variable is within-person productivity improvement from the first two to the middle four rounds. A random-effect specification is used to control for unobserved session effects. The student dummy is coded as 0 for workers and 1 for students. Standard errors are in parentheses. ***, **, and * denote two-tailed significance levels at 1%, 5% and 10% respectively.

Table 8: Distributions of risk-aversion levels

Risk-Aversion Level	Workers (n=340) Count (Percentage)	Students (n=220) Count (Percentage)
0	40 (11.8%)	11 (5.0%)
1	20 (5.9%)	12 (5.4%)
2	40 (11.8%)	21 (9.5%)
3	23 (6.8%)	17 (7.7%)
4	27 (7.9%)	12 (5.4%)
5	25 (7.4%)	7 (3.2%)
6	19 (5.6%)	10 (4.5%)
7	19 (5.6%)	11 (5.0%)
8	48 (14.1%)	50 (22.6%)
9	79 (23.2%)	69 (31.2)
Mean	5.13	6.12
Standard Deviation	3.25	3.07
Median	5	8
Mode	9	9

Table 9: Determinants of productivity improvement for workers: treatment, risk aversion and their interaction

Treatment	RA Centered at	Constant	Treatment§	RA§	Interaction§
NF:PR (n=120)	0	−0.11 (0.31)	0.40 (0.44)	0.20*** (0.04)	−0.13** (0.06)
	1	0.09 (0.28)	0.27 (0.40)		
	2	0.29 (0.26)	0.15 (0.36)		
	3	0.49 (0.24)***	0.02 (0.34)		
	4	0.69 (0.23)***	−0.12 (0.32)		
	5	0.89 (0.23)***	−0.24 (0.32)		
	6	1.09 (0.23)***	−0.37 (0.32)		
	7	1.29 (0.25)***	−0.50 (0.35)*		
	8	1.49 (0.27)***	−0.63 (0.38)**		
	9	1.69 (0.29)***	−0.76 (0.41)**		
NF:PU (n=120)	0	−0.28 (0.30)	0.57 (0.41)	0.19*** (0.04)	−0.12** (0.03)
	1	−0.08 (0.26)	0.45 (0.37)		
	2	0.11 (0.24)	0.33 (0.33)		
	3	0.30 (0.21)	0.20 (0.30)		
	4	0.49 (0.020)**	0.08 (0.28)		
	5	0.68 (0.20)***	−0.04 (0.28)		
	6	0.88 (0.20)***	−0.17 (0.29)		
	7	1.07 (0.22)***	−0.29 (0.31)		
	8	1.26 (0.24)***	−0.41 (0.34)		
	9	1.46 (0.27)***	−0.54 (0.38)*		
F:PR (n=140)	0	−0.52 (0.24)	0.82 (0.36)	0.23*** (0.04)	−0.16*** (0.06)
	1	−0.29 (0.20)	0.66 (0.31)		
	2	−0.06 (0.18)	0.50 (0.27)		
	3	0.17 (0.21)	0.34 (0.14)		
	4	0.39 (0.14)***	0.18 (0.21)		
	5	0.63 (0.14)***	0.02 (0.20)		
	6	0.85 (0.14)***	−0.14 (0.25)		
	7	1.08 (0.15)***	−0.30 (0.10)*		
	8	1.32 (0.18)***	−0.46 (0.28)**		
	9	1.54 (0.20)***	−0.62 (0.32)**		
F:PU (n=140)	0	−0.43 (0.29)	0.73 (0.40)	0.19 *** (0.04)	−0.12** (0.06)
	1	−0.24 (0.25)	0.60 (0.35)		
	2	−0.04 (0.22)	0.48 (0.31)		
	3	0.15 (0.19)	0.35 (0.26)		
	4	0.35 (0.16)**	0.23 (0.23)		
	5	0.54 (0.14)***	0.10 (0.21)		
	6	0.73 (0.14)***	−0.02 (0.22)		
	7	0.93 (0.15)***	−0.15 (0.24)		
	8	1.12 (0.16)***	−0.27 (0.27)		
	9	1.32 (0.19)***	−0.40 (0.32)		

Each regression uses data from the indicated experimental treatment and the control treatment. The dependent variable is productivity improvement from the first two rounds to the middle four rounds. Individual levels of risk aversion are centered as indicated, making the constant and treatment coefficients estimated marginal effects at the indicated risk-aversion level. The treatment dummy variable is coded as 1 for the control and 0 for the treatment. A random-effect specification is used to control for unobserved session effects. Standard errors are in parentheses. § denotes one-tailed tests as explained in the text. ***, **, and * denote significance at 1%, 5% and 10% respectively.

Table 10: Determinants of productivity improvement for students: treatment, risk aversion and their interaction

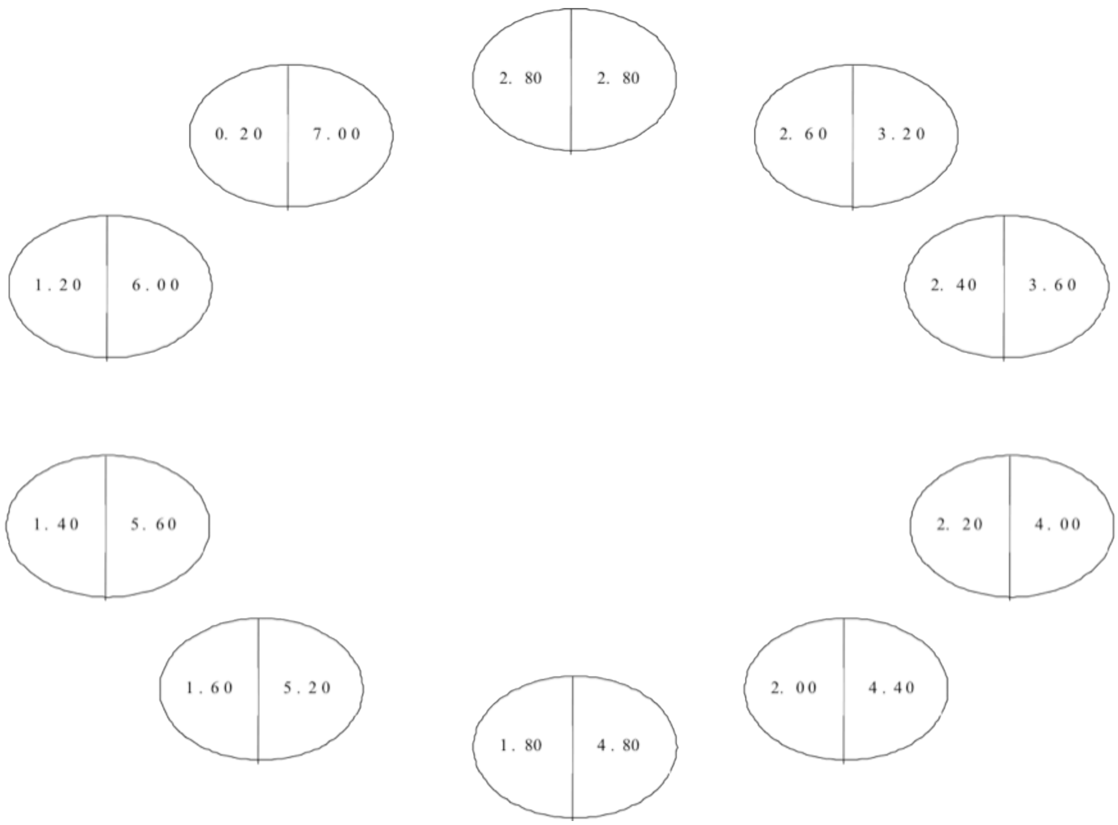
Treatment	RA Centered at	Constant	Treatment§	RA§	Interaction§
NF:PR (n=100)	0	−0.06 (0.55)	1.29 (0.75)	0.23*** (0.07)	−0.34*** (0.11)
	1	0.17 (0.48)	0.95 (0.65)		
	2	0.40 (0.41)	0.61 (0.56)		
	3	0.63 (0.35) **	0.26 (0.48)		
	4	0.86 (0.29) ***	−0.08 (0.40)		
	5	1.09 (0.25) ***	−0.42 (0.35)		
	6	1.32 (0.22) ***	−0.76 (0.33)**		
	7	1.55 (0.21) ***	−1.10 (0.35)***		
	8	1.78 (0.45)***	−1.45 (0.39) ***		
	9	2.01 (0.27)***	−1.79 (0.27) ***		
NF:PU (n=100)	0	0.10 (0.40)	1.13 (0.59)	0.20*** (0.06)	−0.32*** (0.09)
	1	0.31 (0.35)	0.82 (0.51)		
	2	0.51 (0.30) *	0.50 (0.44)		
	3	0.71 (0.25) ***	0.19 (0.37)		
	4	0.91 (0.21) ***	−0.13 (0.32)		
	5	1.1 (0.19) ***	−0.44 (0.29)**		
	6	1.32 (0.18) ***	−0.76 (0.28) ***		
	7	1.52 (0.19) ***	−1.07 (0.29) ***		
	8	1.72 (0.21) ***	−1.39 (0.34) ***		
	9	1.93 (0.25) ***	−1.70 (0.40) ***		
F:PR (n=70)	0	0.51 (0.48)	0.72 (0.65)	0.09 (0.07)	−0.20** (0.10)
	1	0.60 (0.42)	0.52 (0.57)		
	2	0.69 (0.36)*	0.32 (0.49)		
	3	0.78 (0.31)**	0.12 (0.42)		
	4	0.86 (0.28)***	−0.08 (0.37)		
	5	0.95 (0.26)***	−0.28 (0.34)		
	6	1.04 (0.26)***	−0.48 (0.34)*		
	7	1.13 (0.28)***	−0.68 (0.36)**		
	8	1.22 (0.32)***	−0.88 (0.42)**		
	9	1.31(0.37)***	−1.08 (0.49)**		
F:PU (n=70)	0	0.20 (0.68)	1.03 (0.84)	0.28*** (0.10)	−0.39*** (0.12)
	1	0.48 (0.59)	0.64 (0.73)		
	2	0.76 (0.51)	0.25 (0.63)		
	3	1.04 (0.43)***	−0.14 (0.53)		
	4	1.32 (0.36) ***	−0.54 (0.45)		
	5	1.60 (0.31) ***	−0.93 (0.40) ***		
	6	1.88 (0.29) ***	−1.32 (0.38) ***		
	7	2.16 (0.29) ***	−1.71 (0.39) ***		
	8	2.44 (0.45) ***	−2.01 (0.33) ***		
	9	2.72(0.38) ***	−2.50 (0.52) ***		

Each regression uses data from the indicated experimental treatment and the control treatment. The dependent variable is productivity improvement from the first two rounds to the middle four rounds. Individual levels of risk aversion are centered as indicated, making the constant and treatment coefficients estimated marginal effects at the indicated risk-aversion level. The treatment dummy variable is coded as 1 for the control and 0 for the treatment. A random-effect specification is used to control for unobserved session effects. Standard errors are in parentheses. § denotes one-tailed tests as explained in the text. ***, **, and * denote significance at 1%, 5% and 10% respectively.

Figure 1: Sample Workbook

						Answer							Answer
Line 1	69	95	12	72	25		**Line 21**	33	55	40	65	48	
Line 2	95	36	77	85	50		**Line 22**	37	79	88	21	64	
Line 3	80	82	55	24	31		**Line 23**	12	38	12	48	49	
Line 4	65	72	97	87	74		**Line 24**	41	79	33	96	60	
Line 5	25	30	12	72	97		**Line 25**	18	44	68	11	34	
Line 6	83	49	47	37	49		**Line 26**	38	54	83	64	97	
Line 7	30	93	74	71	44		**Line 27**	81	27	85	31	87	
Line 8	87	80	14	17	27		**Line 28**	37	77	21	92	84	
Line 9	51	27	71	76	63		**Line 29**	43	87	83	32	59	
Line 10	31	41	40	10	19		**Line 30**	48	73	94	75	35	
Line 11	15	17	76	46	30		**Line 31**	44	65	79	81	69	
Line 12	68	87	98	49	37		**Line 32**	63	98	72	46	64	
Line 13	14	74	50	85	50		**Line 33**	94	73	54	12	13	
Line 14	25	15	15	10	92		**Line 34**	81	36	43	88	71	
Line 15	20	13	88	22	37		**Line 35**	83	99	38	20	35	
Line 16	59	42	99	50	81		**Line 36**	19	11	99	44	53	
Line 17	48	31	33	15	14		**Line 37**	80	74	91	55	77	
Line 18	56	14	10	77	17		**Line 38**	84	60	55	49	10	
Line 19	60	96	44	33	91		**Line 39**	16	90	41	82	25	
Line 20	86	83	65	47	67		**Line 40**	45	66	14	84	41	

Figure 2: Risk-Aversion Measurement Instrument



Instructions (Translated from the original Chinese)

Your Participant Number is __________.

Thank you for participating today. All of your responses in this study will remain completely anonymous. It is important that during this experiment you do not talk or make any noise that might disrupt others around you. If you have any questions, please raise your hand and the experimenter will answer your questions individually.

During this experiment you will be asked to add up sets of five double-digit integers. There will be nine rounds in which you will be given a number of such sets of five integers. The first round is a trial round for you to get familiar with the task while the remaining eight rounds are experimental rounds, which will be used to calculate your earnings. You are not allowed to use a calculator, but you may write numbers down on scratch paper provided by us. The numbers are randomly drawn and each problem is presented in the following way.

98	42	69	50	78	

You will have a *Workbook* that will contain all of your work. Your job is to solve as many problems as you can in each round. Each round lasts 3 minutes. Your earnings in this experiment will depend on your performance and/or the specific compensation method applied to each of the eight experimental rounds. Once we begin the experiment, you will not be able to look ahead to future pages or to go back to previous pages.

To ensure anonymity, please write down only your participant number on each page of the *Workbook*. Please do not write your name on any of these materials. The data will only be identified by the participant code assigned to you and will not at any point be connected to your name or face in any way. Please make sure that you completely understand the instructions for the experiment. Once again, remember not to make any noises that might disturb others around you. If you have any questions, raise your hand and we will answer your questions individually.

C: You will earn RMB 6.00 in this round, regardless of the number of words you create in this round.

NF:PR: You will earn RMB 6.00 in this round, regardless of the number of words you create in this round. In addition, your performance will be ranked against everyone else in the session and your ranking information will be reported to you privately.

NF:PU: You will earn RMB 6.00 in this round, regardless of the number of words you create in this round. In addition, your performance will be ranked against everyone else in the session and everyone's ranking information will be linked to his/her name and will be reported to everyone publicly.

F:PR: Your performance will be ranked against everyone else in the session and your ranking information will be reported to you privately. You will be paid based on your performance ranking, F:PU: Your performance will be ranked against everyone else in the session and everyone's ranking information will be linked to his/her name and will be reported to everyone publicly. You will be paid based on your performance ranking.