EES 2004 : Experiments in Economic Sciences - New Approaches to Solving Real-world Problems

September 27, 2004 **Preliminary and Incomplete**

The Effects of Compensation Schemes on Self-Selection and Work Productivity: An Experimental Investigation

C. Bram Cadsby*
Department of Economics
University of Guelph

Fei Song
Schulich School of Business
York University

Francis Tapon
Department of Economics
University of Guelph

Abstract

This research examines experimentally the impact of productivity-based versus lump-sum compensation schemes on self-selection and work performance. We find that as predicted by Jensen (2003), participants in a laboratory experiment with salient incentives self-select themselves into preferred compensation schemes based at least partially on performance. However, the compensation scheme selected is also based on individual attitudes toward risk. Individuals demonstrating a higher degree of risk-aversion in a lottery-selection task exhibit a higher probability of selecting the risk-free lump-sum compensation scheme, while less risk-averse individuals are more likely to select the productivity-based scheme. A laboratory firm offering linear compensation achieved significantly higher productivity than an identical firm offering flat compensation for two reasons: first, more highly skilled workers selected it, and second, workers on average, regardless of their self-selections, were more productive under the linear scheme.

*We would like to thank the Economics Department at the Queensland University of Technology for their warm hospitality and support for this research. Cadsby would also like to acknowledge generous funding from the Social Sciences and Humanities Research Council of Canada, grant #410-2001-1590.

1. Introduction

This research examines experimentally the impact of productivity-based versus lump-sum compensation schemes on self-selection and work performance. Michael Jensen (2003) argues that the shape of a firm's compensation scheme, i.e. whether people are paid a lump-sum or on the basis of performance, leads them to self-select into a preferred compensation scheme based on their productivity. In particular, he argues that more productive workers will choose the productivity-based compensation scheme, while less productive workers will opt for the lump-sum scheme. In addition, he argues that once a choice is made, the nature of the compensation scheme affects productivity with the productivity-based scheme motivating greater productivity than the lump-sum scheme. Although the latter claim has been examined extensively theoretically (Baker, Jensen, & Murphy, 1988; Lazear, 2000; O'Dell, McAdams, 1987) as well as empirically in both the laboratory (e.g., Fessler, 2003; Kuhn & Yockey, 2003; see also Camerer & Hogarth, 1999 and Bonner, Hastie, Sprinkle, & Young, 2000 for reviews of the literature) and the field (e.g., Abowd, 1990; Stajkovic & Luthans, 2001), the former has not. However, Jensen maintains that for both of these reasons, a productivity-based compensation scheme should lead to a higher level of productivity than a lump-sum scheme.

Jensen's arguments provoke a number of questions. First, in order to self-select themselves as Jensen predicts, people must formulate forecasts or expectations about their own future productivity. To the extent that people are overly self-confident and optimistic (Taylor, 1989; Taylor & Brown, 1988), they may self-select themselves into the productivity-based scheme when they would have done better under the lump-sum scheme. The "better-than-average" effect reported by a number of behavioral studies suggests that such optimism might be prevalent among people who place a high value on their productivity at work, but have little concrete evidence on how productive they actually are (Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995; Epley & Dunning, 2000; Miller & Ratner, 1998). For example, Camerer and Lovallo (1999) explore whether optimistic biases could plausibly and predictably influence economic behavior in one particular setting: entry into competitive games or markets. In particular, they examine and test a possible explanation for business failure as a result of managers acting on the basis of excessive optimism about the relative skills they possess. The findings are consistent with the prediction that overconfidence leads to excessive business entry. Similarly, to the extent that people are lacking self-confidence and overly pessimistic, they may self-select themselves into the lump-sum scheme when they would have done better under the productivity-based scheme. To what extent do people choose the scheme that maximizes their income? Do people improve their predictions and hence their choice of compensation scheme with experience? Are expectations rational enough to corroborate the proposition relating self-selection to actual productivity?

Second, since the lump-sum scheme offers a certain outcome while the productivity-based scheme involves uncertainty, standard economic theory would predict that in addition to expected earnings from the different compensation schemes, the level of risk-aversion should also affect scheme selection. What role does risk-aversion play in the selection of a compensation scheme? Should organizations with productivity-based schemes expect their work forces to be less risk-averse than those with lump-sum schemes?

Third, do demographic factors such as gender or native language (the experimental task involves a knowledge of English words) affect self-selection into a compensation scheme either through their possible correlation with productivity and/or risk-aversion, or independently of these factors? Gender has been examined in the risk preference literature but the results are somewhat mixed and inconclusive (e.g., Dufwenberg & Gneezy, 2003; Eckel & Grossman, 2000; Shupp & Williams, 2003). It is thus interesting to look at the gender effect in this specific context where risk preference may play a role in self-selection.

In regard to native language, we follow Camerer and Hogarth's (1999) theoretical framework and consider native language as a person's cognitive capital in this experimental task. Specifically, Camerer and Hogarth (1999) argue that an experiment is a cognitive activity in which subjects participate in a production process involving both capital and labor. Capital is the knowledge a subject has, labor is the effort a subject puts into the experimental task, and production is the outcome of the task. Hence, as experimental economists, we are interested in "how well can subjects maximize their objective function, given available capital and a particular production function" (Camerer & Hogarth, 1999, p. 9). Consequently, it is important to look at the effect of cognitive capital, an important component of which is native language, in the case of our experiment. This framework suggests that it affects self-selection only insofar as it affects productivity.

Fourth, are people more productive in performing a complex task under the productivity-based versus the lump-sum scheme regardless of their revealed preference for one scheme over the other?[1] In particular, do those selecting the stress-free safety of the lump-sum scheme actually perform better without the stressful financial incentives of the productivity-based scheme? If not, do financial incentives increase their productivity by as much as such incentives increase the productivity of those who prefer the productivity-based scheme?

Fifth, to what extent is higher productivity under the productivity-based scheme due to the self-selection of more productive workers into that scheme and to what extent is it due to a reduction of effort when financial incentives are lacking?

2. Experimental Design

Participants were recruited at a large, urban Australian university by means of both announcements in economics classes and random recruitment in the lounge area of the business school. All 115 participants were undergraduate students and most, but not all, were majors in economics or other subjects taught within the business faculty.

Upon arrival, the experimental instructions were read to the participants while they followed along on their own copies. A copy of the instructions may be found in Appendix 1. Participants were asked to play one practice and eight experimental three-minute word-creation (anagram) games using prescribed sets of seven letters. They were provided with a prepared workbook in which to perform this task. Each anagram was presented on a separate page of the workbook. Other pages were used for participants to record their choices of compensation scheme or devoted to explaining which compensation scheme would apply in a subsequent anagram round. Participants were

[1] A complex task is defined as one requiring individuals to determine ways to achieve a certain goal when solutions are unclear (Bonner, et al., 2000). We employ an anagram task, which falls into this category.

not permitted to look ahead to future pages or to go back to previous pages. They were allowed to tear off one page and look at the next only when instructed to do so by the experimenter. To ensure anonymity, players wrote their assigned participant numbers, but not their names on each page of the workbook immediately prior to beginning work on that page.

The experiment utilized two different compensation schemes. The first compensation scheme paid $0.20 per correct word created. We will henceforth refer to this linear productivity-based scheme as the "linear" scheme. The second scheme paid a lump sum of $2.20, independent of performance. We will henceforth refer to this flat-rate scheme as the "flat" scheme. Since a person creating 11 words under the productivity-based scheme would earn $2.20, 11 words was the break-even point between the two schemes.

Table 1 presents the anagrams used by round along with the average productivity for each anagram in Vance and Colella (1990) and in a pre-test we ran prior to the current study. The Vance and Colella (1990) study used psychology undergraduates from Ohio State University. Participants were given performance targets, but were not paid on the basis of performance. Our pre-test employed 99 business undergraduates at a large Canadian business school who were given salient performance-based financial incentives. The average productivity in our pre-test was higher than in the Vance and Colella (1990) study in eight of the nine rounds. However, the differences from round to round were similar. The overall mean excluding the practice round was 9.65 words in Vance and Colella (1990) and 12.57 in our pre-test. Since the current study involves both linear and flat compensation schemes, we chose the average of these two numbers, approximately 11 words, to set as our break-even point.

At the beginning of the experiment, participants chose which one of the two compensation schemes they would like to adopt for calculating their earnings for rounds 1 and 2. After playing a practice round, they played rounds 1 and 2 and were paid according to the scheme they had selected. For rounds 3 and 5, all participants were paid according to the flat scheme regardless of their earlier choice. For rounds 4 and 6, all participants were paid according to the linear scheme, regardless of their earlier choice. In each case, they were informed of the payment scheme immediately prior to the round. For rounds 7 and 8, participants were again given the choice between the two compensation schemes. After each round, each participant's list of words was collected by the experimenters and taken to another room were the number of correct words was calculated. Participants did not receive feedback on the number of correct words they had produced until they were paid at the end of the session.

This design enabled us to examine how the self-selection decision evolved with experience. It also allowed us to compare the performance of those who selected the linear scheme with those who selected the flat scheme under their chosen schemes. In addition, it permitted us to compare the productivity of each self-selected group under each of the two schemes, and to examine whether these groups reacted differently to the introduction of financial incentives.

Afterwards, participants completed a questionnaire, in which they responded to a number of demographic questions. However, the primary purpose of the questionnaire was to elicit risk preferences. This was accomplished by asking participants to make ten lottery-choice decisions based on an instrument developed by Holt and Laury (2002). The questionnaire containing the lottery choices is attached as Appendix 2. One of the paired lottery choices was randomly selected and implemented in each session. In addition to being paid for the words they created according to the compensation schemes outlined above, participants were paid an additional sum based on the outcome of their chosen lottery from the pair of randomly-selected lotteries.

The purpose of eliciting risk preferences was to examine the role of such preferences in the self-selection of participants into payment schemes. Holt and Laury (2002) found that risk preferences were affected by the amount of money at stake. In particular, larger stakes were associated with a higher level of risk aversion. We therefore adjusted the stakes used by Holt and Laury (2002) to correspond as closely as possible to the amount at stake in the two rounds of the anagram game affected by each self-selection decision. This involved multiplying Holt and Laury's (2002) lottery numbers by 2.2 to obtain the appropriate amounts in Australian dollars. At the end of the session, players were taken individually to another room where they were paid privately in cash. On average, participants earned $21.20 AUD (about $15.11 US) for a session lasting approximately one hour and 15 minutes.

3. Results

All 115 participants completed the study. For rounds 1 and 2, 57 participants chose the linear scheme, while 58 selected the flat scheme. For rounds 7 and 8, 60 chose the linear scheme, while 55 chose the flat scheme.

We first examine the percentage of participants who gained, broke even and lost under their selected compensation scheme versus the alternative scheme. Table 2 reports the results. Recall that 11 words is the break-even point. For those who chose the linear scheme, 52.8% produced an average of less than 11 words in rounds 1 and 2, thus making less than they would have under the flat scheme. This suggests that over half of those selecting the linear scheme prior to playing the game were overly optimistic, consistent with the better-than-average effect. In contrast, only 22.4% of those who chose the flat scheme produced an average of more than 11 words, thus making less than they would have if they had produced the same amount under the linear scheme. This may have been because pessimism about one's own ability was less prevalent than optimism or because some of those who were able to produce more than 11 words with effort were not motivated to exert such effort under the flat scheme. To distinguish between these two possibilities, we also calculated the productivity of those who initially selected the flat scheme in rounds 4 and 6 when they were compelled to produce under the linear scheme. Increased effort made only a small difference. Only two more participants, resulting in a total of 25.9%, made more money under the linear scheme than they had made under the flat scheme in rounds 1 and 2. It would thus seem that a bias towards overly-optimistic self-confidence was primarily responsible for the higher percentage of participants who made the wrong choice, financially speaking, at the first self-selection.

By the beginning of round 7, participants had gained experience with the game. Although they had not received explicit feedback on their performance from the experimenters, they knew how many words they had submitted for each anagram and presumably had a reasonable idea of how many were likely to be correct. Thus, we might expect that they would have more realistic expectations of their subsequent performance. This indeed appears to be the case for those who chose the linear scheme for rounds 7 and 8, only 28.3% of whom made less money than they would have done under the flat scheme compared to 52.6% for rounds 1 and 2. Thus, the excessive unrealistic optimism exhibited initially seems to decline with experience. This experience effect was much less dramatic for those choosing the flat scheme for rounds 7 and 8, 18.2% of whom made less money than they would have done had they produced the same number of words under the linear scheme, compared to 22.4% for rounds 1 and 2. Applying the rounds 7 and 8 selection criteria to rounds 4 and 6 when everyone played under the linear scheme, the corresponding number was 14.5%.

Was the probability of at least breaking even for those choosing the linear scheme significantly greater than the probability of at least breaking even for those choosing the flat scheme? In the former case, this implies producing 11 words or more, while in the latter case it implies producing 11 words or less. We examine this question by running the following logistic regression separately for the initial and final self-selection cases:

$$\ln [p_e/(1-p_e)] = \beta_0 + \beta_1 \cdot SS_i \qquad (1)$$

where p_e is the probability of earning at least as much under the self-selected compensation scheme as under the alternative scheme and SS_i is a dummy variable equal to 1 if the linear scheme is selected and 0 if the flat scheme is selected. The i subscript refers to the initial and final selections. Table 3 reports the results. Initially, those selecting the linear scheme exhibit a significantly lower probability of at least breaking even than those selecting the flat scheme ($p = 0.000$). However, this significance disappears with the final self-selection. Thus, the apparent bias towards more rather than less self-confidence vanishes with experience, consistent with evidence indicating that the better-than-average effect disappears with performance feedback.

To examine how participants learn from experience, we next compare the mean productivity in rounds 3, 4, 5, and 6 for those who maintained and those who altered their choice of compensation scheme. The pairwise tests reported in Table 4 indicate a significant difference between those who maintained a choice of the flat scheme through both self-selection decisions and those who changed their choice from flat to linear ($p = 0.001$). Similarly, there was a significant difference between those who maintained a choice of the linear scheme and those who changed their choice from linear to flat ($p = 0.001$). Thus, despite not receiving explicit feedback on the number of correct words they had produced during the game, participants seem to have a good idea of how well they are doing and to respond accordingly.

In order to assess simultaneously the role of attitude toward risk and productivity in the selection of compensation scheme, we elicited risk preferences using an instrument developed by Holt and Laury (2002) as described in the discussion of the experimental design above. The ten lottery decisions presented to the participants (see Appendix 2 for details) each involve a relatively safe choice (option A) versus a relatively risky choice (option B). The probabilities of each lottery outcome are manipulated so that each decision involves progressively higher expected earnings for the risky choice relative to the safe choice. Accordingly, everyone should have a switching point, above which safe choices are selected and below which risky choices are selected. Of the 115 participants, 99 had in fact one switching point, consistent with expected utility theory. In our analysis, involving attitudes toward risk, we followed the cautious approach of discarding data from those who exhibited more than one switching point, leaving us with 99 usable data points. The 99 choices made are summarized in Table 5, which follows the format of Holt and Laury's (2002) Table 3.

Our participants were highly risk-averse with 93% exhibiting some degree of risk-aversion. Of the others, 3.5% were risk-neutral, while another 3.5% were risk-loving. These levels of risk-aversion are somewhat higher than those found by Holt and Laury (2002) in their lower stakes setting and roughly comparable to those found in their higher stakes setting.

We are now able to investigate the roles played by risk-aversion and productivity on the compensation scheme selected, both for the initial two rounds and for the final two rounds, utilizing a logistic regression with a simple two-by-two factorial design as follows:

$$\ln [p_{ssi}/(1-p_{ssi})] = \beta_0 + \beta_1 \cdot \text{Productivity} + \beta_2 \cdot \text{Risk-Aversion} + \beta_3 \cdot \text{Interaction} \qquad (2)$$

where p_{ssi} = the probability of self-selecting the linear scheme and i refers to the initial versus the final selections. A separate estimation is run for each of the two selection decisions.

The first factor is each participant's level of risk-aversion as measured by the lottery mechanism. The null hypothesis is contrasted with the alternative suggested by theory that higher levels of risk aversion are associated with a lower probability of selecting the linear scheme. The second factor is the productivity of each participant as measured by the data from the four middle rounds when all players were compensated in the same manner. The null hypothesis is contrasted with the alternative suggested by Jensen (2003) that higher levels of productivity are associated with a higher probability of selecting the linear scheme. When expected earnings are greater under the flat scheme as they would be if a participant expected to produce fewer than 11 words, theory predicts that only those who are risk-loving would choose the linear scheme. Since only 3.5% of participants are risk-loving, we would expect that risk attitudes would play little role when productivity is low and a far more substantial role when productivity is higher. We thus test the null hypothesis on the interaction of risk-attitude and productivity against the alternative that higher levels of productivity are associated with a more negative impact of the level of risk-aversion, resulting in a predicted negative interaction effect. Accordingly, we center productivity at the break-even point of 11 words so that the coefficient on risk-aversion is estimated at the lowest point where it is likely to have a significant impact. Risk-aversion is more conventionally centered at its mean so that the coefficient on productivity is estimated at the mean level of risk-aversion.

Table 6 reports the results. For the initial self-selection, the null hypotheses for both productivity ($p = 0.003$) and risk-aversion (0.040) are both rejected in the direction of the specified alternatives. However, the interaction is not significant ($p = 0.123$). Thus, despite the excessive amounts of self-confidence noted above, those who subsequently perform better do have a significantly higher probability of self-selecting themselves into the linear scheme, controlling for their differing attitudes toward risk. At the same time, controlling for productivity, those who exhibit a higher degree of risk-aversion in their lottery choices are more likely to choose the risk-free flat compensation scheme. The lack of significance of the interaction effect suggests that the expected earnings of most participants were higher under the linear than under the flat scheme so that risk-aversion mattered significantly and in a quantitatively similar manner even for those who expected to perform relatively poorly. This is consistent with the excessive self-confidence documented earlier.

For the final self-selection, the null hypotheses for both productivity ($p = 0.000$) and risk-aversion (0.0150) are also both rejected in the direction of the specified alternatives. The coefficients are larger and the p-values smaller than in the initial selection. The null hypothesis for the interaction is also rejected in the direction of the predicted negative effect ($p = 0.048$). This is consistent with more participants expecting to produce fewer than 11 words than under the initial selection. Risk-aversion matters less for those producing less because many of them do not expect higher earnings under the risky linear scheme and hence opt for the flat scheme regardless of their degree of risk aversion. Clearly, self-selection into the two compensation schemes depends on both productivity and risk-aversion in the manner predicted.

We also examined the role played by gender. Of our participants, 44 were female and 71 were male. We began by examining whether or not gender was related to the two independent variables in our study: productivity and risk-aversion. Table 7 reports the results. Females were significantly more productive than males ($p = 0.045$) when productivity is measured across both the flat and linear middle rounds. When measured using only the linear rounds, this difference in productivity was not significant ($p = 0.120$), though it was significant in the flat rounds ($p = 0.026$). The mean levels of risk-aversion were almost identical for males and females and thus not significant. We then added gender to the logistic regression to test whether or not it had any effect on the self-selection decision apart from its effect on productivity. It had no such effect.

Similarly, we examined the role played by native language. Of our participants, 85 were native speakers of English, while 30 were not. Since success at producing English words from an anagram is related to knowledge of English, we would expect to find a strong relationship between native language and productivity. The results reported in Table 8 confirm this is the case whether productivity is measured across both the flat and linear rounds ($p = 0.003$) or for the linear ($p = 0.001$) or flat (0.023) rounds alone. The mean levels of risk-aversion are virtually identical and hence their difference is insignificant. Adding a native language dummy to the logistic regression produced insignificant results, indicating that language matters only insofar as it affects productivity.

Do people generally perform better under the linear than under the flat compensation scheme? In rounds 3 and 5, everyone is compensated using the flat scheme, while in rounds 4 and 6, everyone is compensated using the linear scheme. We are thus able to perform a within-person comparison of productivity under the two schemes. The results, reported in Table 9, indicate that as expected, participants perform significantly better under the linear than under the flat compensation scheme ($p = 0.000$). Note that this is the case even though the mean levels of productivity were slightly higher for the anagrams used in the flat compensation scheme than for those used in the linear scheme. There is however another possible confounding factor. Participants may improve with practice. The two flat rounds 3 and 5 are run prior to the two linear rounds 4 and 6. To remove this confound, we compared productivity in the earlier linear round 4 with that in the later flat round 5. Although the difference was lower and the p-value higher, the linear scheme still resulted in significantly higher productivity than the flat scheme ($p = 0.042$).

We have now established that people who self-select themselves into the linear scheme perform better than those who self select themselves into the flat scheme for two reasons. First, people with more skill at the task are more likely to select the linear performance-based compensation scheme. Second, people perform better on average under the linear than under the flat scheme. In figure 1, we decompose the higher productivity of those choosing the linear over those choosing the flat scheme into these two components. The left half of the figure deals with the initial self-selection for rounds 1 and 2, SS1. The right half of the figure deals with the final self-selection for rounds 7 and 8, SS2. Each productivity number represents the mean level of productivity for those working under a particular compensation scheme in the indicated rounds, given the compensation scheme selected. L_L represents those working under the linear scheme who also chose the linear scheme; F_F represents those working under the flat scheme who also chose the flat scheme; L_F represents those working under the linear scheme who chose the flat scheme; and F_L represents those working under the flat scheme who chose the linear scheme. Differences are indicated between the arrows and p-values associated with those differences are presented in parentheses.

The numbers on the left under the header 1,2 represent data from rounds 1 and 2. Those selecting and hence working under the linear scheme are significantly more productive than those selecting and hence working under the flat scheme ($p = 0.013$). To decompose this difference, it is necessary to examine the productivity of those selecting each scheme in rounds 3 and 5, in which everyone works under the flat scheme, versus rounds 4 and 6, in which everyone works under the linear scheme. For those selecting the linear scheme initially, average productivity in rounds 4 and 6 under the linear scheme is 11.43 words. For those selecting the flat scheme initially, average productivity in rounds 3 and 5 under the flat scheme is 8.71 words. The difference of 2.72 words is significant (p = 0.000) just as the comparable difference was significant in rounds 1 and 2. This difference can be decomposed into one component due to “skill” differentials between persons making different self-selections working under a common scheme and another due to the “incentive” effects of working under the different schemes. The skill differential under the linear scheme is 1.73 words (p = 0.001), while the incentive effect for those selecting the flat scheme is 0.99 words (p = 0.000). The incentive effect for those selecting the linear scheme is 1.26 words (p = 0.000), while the skill differential under the flat scheme is 1.46 (p = 0.006). Thus, both skill differences between those self-selecting into the different schemes and motivational differences between the two schemes contribute significantly to productivity differences between workers who are allowed to self-select into the compensation scheme of their choice.

The numbers on the right under the header 7,8 represent data from rounds 7 and 8. Comparing the mean productivity of those selecting the linear scheme in these rounds to those selecting it in rounds 1 and 2 reveals a significant increase of 1.89 words. The slight decrease in the average productivity of those selecting the flat scheme is not significant (p = 0.533). However, the differential between those selecting and working under the linear scheme and those selecting and working under the flat scheme has grown to 3.52 words, which continues to be significant (p = 0.000). When the data in the middle rounds is resorted based on the final self-selection decision, there is little change in the motivational differences. However, the resorting increases the skill differentials, indicating that participants have made self-selections more in tune with their actual abilities once having experience with the experimental task.

Any notion that those selecting the flat compensation scheme might perform better in that relatively stress-free environment is dispelled by the positive and significant motivational differentials under both schemes whether based on the initial or final self-selections. The pairwise tests reported at the bottom of the figure indicate further that there is no significant difference between the motivational effect of pay-for-performance based on self-selection ($p = 0.452$ for the first self-selection, $p = 0.601$ for the second self-selection).

It is interesting to note that 29 out of 115 participants, a substantial minority, did perform better under the flat than under the linear scheme. Of these, 12 chose flat and 17 chose linear in the initial self-selection, while 14 chose flat and 15 chose linear in the final self-selection. Thus, there is no relationship between the scheme chosen and the scheme motivating higher productivity for a given individual. Of course, most individuals are likely more concerned with weighing their expected earnings under each scheme against the uncertainty associated with the linear scheme than with determining under which scheme they would likely produce the most for their employers. Thus, in more than 50% of observed cases, people performing better under the flat scheme self-selected into the linear scheme where their risk-adjusted expected returns were presumably higher.

4. Conclusion and Managerial Implications

At least since the pioneering paper by Jensen and Meckling (1976), economists and management scholars have been acutely aware of the importance of the relationship between performance and incentives. Brickley, Smith and Zimmerman (2004) liken the allocation of decision rights, performance incentives and the standards used in performance evaluation to the three legs of a three-legged stool enabling a firm's organizational architecture to stay in balance and the firm to operate effectively. In this paper, we focus on the role of performance incentives, concentrating on the links between linear productivity-based versus flat lump-sum compensation schemes, performance and attitudes towards risk.

We find that as predicted by Jensen (2003), participants in a laboratory experiment with salient incentives do self-select themselves into preferred compensation schemes based at least partially on performance. This occurs because performance expectations are directly related to expected earnings under the two available schemes. Although performance expectations initially show an overly optimistic bias consistent with the widely-reported better-than-average effect and the recent experimental paper by Camerer and Lovallo (1999), this bias disappears with experience.

However, the compensation scheme selected is also based on individual attitudes toward risk. Individuals demonstrating a higher degree of risk-aversion in a lottery-selection task exhibit a higher probability of selecting the flat risk-free compensation scheme, while less risk-averse individuals are more likely to select the linear scheme.

A laboratory "firm" offering linear compensation achieved significantly higher productivity than an identical firm offering flat compensation for two reasons: first, more highly skilled workers selected it, and second, workers on average, regardless of their self-selections, were more productive under the linear scheme. Although a substantial minority of workers were more productive under the flat than under the linear scheme, they were no more likely to select the flat scheme than members of the majority who performed better under the linear scheme.

Our analysis offers some important lessons for managers. First, employees will self-select themselves among companies depending on the compensation scheme offered with relatively less productive employees gravitating toward companies offering a flat salary regardless of performance, while the most productive employees join firms where much of the compensation is performance-based. This appears to suggest that performance-based pay schemes should be prevalent throughout industry. However, there are many situations in which it is difficult to implement such schemes. For example, it is not always easy to observe the productivity of individual employees. Sometimes it is very costly to do so. Allowing employees to self-report can lead to considerable deception, which can have devastating consequences for organizations (Jensen, 2003). This kind of situation seems to require non-performance based salary compensation. In such cases, special attention must be paid to motivating high levels of performance through means other than compensation.

In other cases, several employees' jobs are inter-related and it is impossible to disentangle the contributions of specific individuals. Examples of such jobs abound in the automotive industry in which teams cooperate to design new models. Each specialty relies on every other and depends on cooperation to get the job done. In such cases, compensation may be based on team productivity or based on a salary independent of performance. Further studies need to be done to assess the effects of such team compensation schemes on self-selection into team-related occupations.

Second, employees who self select themselves into performance-based compensation schemes may be more productive but also tend to be less risk-averse, and as a result possibly more prone to bending or even breaking rules. This may create enormous problems for their firms, possibly leading all the way to bankruptcy. One example is the bond trading scandal at Salomon Brothers in the early 1990s, which almost led to the firm being shut down by regulators. Jensen (2003) discusses at length how compensation systems with bonuses linked to the attainment of performance thresholds are a guarantee of illegal and/or value- destroying behavior. Examples abound in the pharmaceutical and consumer goods industries. Our experiments also show that performance-based compensation schemes attract less risk-averse individuals who, although more productive, may also require more vigilance on the part of senior management and company directors. The cost of such vigilance must be weighed against the productivity benefits of performance –based compensation.

References

Abowd, J. (1990). “Does Perfermance-Based Managerial Compensation Affect Corporate Performance,” *Industrial and Labor Relations Review*, 43, 52-74.

Alicke, M. D., & Klotz, M. L. (1995). “Personal contact, individuation, and the better-than-average effect,” *Journal of Personality and Social Psychology*, 68, 805-825.

Baker, G. Jensen, M. & Murphy, K. (1988). “Compensation and Incentives: Practice versus Theory,” *Journal of Finance*, 43, 593-617.

Bonner, S. E., Hastie, R., Sprinkle, G. B., & Young, M. (2000). “A Review of the Effects of Financial Incentives on Performance in Laboratory Tasks: Implications for Management Accounting,” *Journal of Management Accounting Research*, 12, 19-57.

Brickley, J.A., C.W.Smith C.W. Jr., & Zimmerman, J.L. (2004). *Managerial Economics and Organizational Architecture. 3rd Ed.* (Irwin).

Camerer, C.; & Hogarth, R. M. (1999). “The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production,” *Journal of Risk and Uncertainty*, 19, 7-42.

Dufwenberg, M., Gneezy, U., & Rustichini, A. (2004). "Price Competition: The Role of Gender and Education". Working Paper.

Eckel, C., & Grossman, P. (2000) "Men, Women, and Risk Aversion". In Plott and Smith (Eds.), Handbook of Results in Experimental Economics.

Epley, N., & Dunning, D. (2000). “Feeling ‘holier than thou’ Are self-serving assessments produced by errors in self- or social prediction?” *Journal of Personality and Social Psychology*, 79, 861-875.

Fessler, N. J. (2003). "Experimental Evidence on the Links among Monetary Incentives, Task Attractiveness, and Task Performance," *Journal of Management Accounting Research*, 15, 161-182.

Holt, C. & Laury, S. (2002)."Risk Aversion and Incentive Effects," *American Economic Review*, 92, 1644-1655.

Jensen, M. C. (2003). "Paying People to Lie: The Truth about the Budgeting Process," *European Financial Management*, 9, 379-406.

Jensen, M. C. & Meckling, W. H. (1976). "Theory of the firm: Managerial behavior, agency costs and ownership structure," *Journal of Financial Economics* 3, 305-360

Kuhn, K. M.; & Yockey, M. D. (2003). "Variable Pay as a Risky Choices: Determinants of the Relative Attractiveness of Incentive Plans," *Organizational Behavior and Human Decision Processes,* 90, 323-345.

Lazear, E. P. (2000). "The Power of Incentives," *American Economic Review*, 90, 410-415.

Miller, D. T., & Ratner, R. K. (1998). The disparity between the actual and assumed power of self-interest. *Journal of Personality and Social Psychology*, 74, 53-62.

O'Dell, C.; & McAdams, J. (1987). "The Revolution in Employee Rewards," *Management Review*, 76, 30-34.

Shupp, R. S., & Williams, A. (2003). "Risk Preference Differentials of Small Groups and Individuals". Working Paper.

Stajkovic, A. D., &Luthans, F. (2001). "Differential effects of incentive motivators on work performance," *Academy of Management Journal*, 44, 580-601.

Taylor, S. E. (1989). *Positive Illusions: Creative Self-Deception and the Healthy Mind.* New York, NY, US: Basic Books, Inc.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin,* 103 (2), 193-210.

Table 1
Anagrams used in the Experiment

ROUND	ANAGRAM	VANCE AND COLELLA (1990) MEAN	PRE-TEST MEAN
Practice	OASFKEV	8.63	8.56
Round 1	OADMHUP	11.16	14.58
Round 2	AEDBKUG	8.95	10.98
Round 3	0ELBJAM	9.42	12.15
Round 4	UADQWER	8.84	11.89
Round 5	EASCKIY	9.32	12.09
Round 6	AODJGIP	9.63	12.20
Round 7	UONHMEY	10.21	13.06
Round 8	OELHMAZ	9.63	13.62
Round 1 & 2		10.06	12.78
Round 3 & 5		9.37	12.12
Round 4 & 6		9.24	12.05
Round 7 & 8		9.92	13.34
All rounds exc. Practice		9.65	12.57

Table 2
Number of People who Gained, Broke Even, and Lost under each Compensation Scheme

Panel A: Productivity-Based (Linear) Compensation Scheme

	Rounds 1&2	**Rounds 7&8**
>11 (gained)	22 (38.6%)	39 (65%)
=11 (broke even)	5 (8.8%)	4 (6.7%)
<11 (lost)	30 (52.6%)	17 (28.3%)
Total	57	60

Panel B: Lump-Sum (Flat) Compensation Scheme

	Rounds 1&2	**Rounds 7&8**
>11 (lost)	13 (22.4%)	10 (18.2%)
=11 (broke even)	2 (3.4%)	4 (7.3%)
<11 (gained)	43 (74.1%)	41 (74.5%)
Total	58	55

Panel C: L Productivity in Rounds 4 and 6 for Those Who Chose Lump-Sum (Flat) Compensation Scheme Initially and Finally

	Chose Flat Scheme Initially	**Chose Flat Scheme Finally**
>11 (lost)	15 (25.9%)	8 (14.5%)
=11 (broke even)	2 (3.4%)	4 (7.3%)
<11 (gained)	41 (70.7%)	43 (78.2%)
Total	58	55

Table 3
Mean Productivity and Pairwise Tests for those who Maintained and those who Altered their Choice of Compensation Scheme

Panel A: Means

(INITIAL, FINAL)	PRODUCTIVITY IN ROUNDS 3, 4, 5, 6	N
(Flat, Flat)	8.47	39
(Flat, Linear)	10.71	19
(Linear, Linear)	11.51	41
(Linear, Flat)	8.97	16

Panel B: Pairwise Tests

COMPARISON	DIFFERENCE	T-STAT.	P-VALUE
(Flat, Flat) vs. (Flat, Linear)	-2.24	-3.34	0.001
(Linear, Linear) vs. (Linear, Flat)	2.54	3.68	0.001

Table 4
Logistic Regression of Probability of Breaking Even or Better on Self-Selection of Compensation Scheme (two-tailed P-values in parentheses)

INDEPENDENT VARIABLE	INITIAL SELF-SELECTION (ROUNDS 1 AND 2)	FINAL SELF-SELECTION (ROUNDS 7 AND 8)
Constant	1.242 (.001)	1.504 (.000)
Self-Selection (Flat=0, Linear=1)	-1.347 (.000)	-.576 (.203)

Table 5
Risk-Aversion Classifications Based On Holt-Laury Lottery Choices

Number of safe choices	Range of relative risk aversion for $U(x) = x^{1-r} / (1-r)$	Risk preference classifications	Proportion of choices
0-1	$r < -0.95$	highly risk loving	3.5%
2	$-0.95 < r < -0.49$	very risk loving	0%
3	$-0.49 < r < -0.15$	risk loving	0%
4	$-0.15 < r < 0.15$	risk neutral	3.5%
5	$0.15 < r < 0.41$	slightly risk averse	9.6%
6	$0.41 < r < 0.68$	risk averse	25.2%
7	$0.68 < r < 0.97$	very risk averse	22.6%
8	$0.97 < r < 1.37$	highly risk averse	21.7%
9-10	$1.37 < r$	stay in bed	13.9%

Table 6
Logistic Regression of the Probability of Choosing the Linear Productivity – Based Compensation Scheme as a Function of Actual Productivity and Attitude Toward Risk (two-tailed p-values for constants and one-tailed p-values for explanatory variables in parentheses)

INDEPENDENT VARIABLE	INITIAL SELF-SELECTION	FINAL SELF-SELECTION
Constant	0.280 (0.226)	0.921 (0.004)
Productivity (Centered at 11)	0.241 (0.003)	0.574 (0.000)
Risk-Aversion (Centered at Mean)	-0.264 (0.040)	-0.450 (0.015)
Interaction	-0.064 (0.123)	-0.136 (0.048)

Table 7
Role of Gender

	FEMALE (N=44)	MALE (N=71)	DIFFERENCE	P-VALUE
Productivity (3,4,5,6)	10.64	9.60	1.04	0.045
Linear Productivity (4,6)	11.09	10.23	0.86	0.120
Flat Productivity (3,5)	10.18	8.96	1.22	0.026
Risk Aversion*	6.92	6.86	0.06	0.879

*Of those whose responses were consistent, 63 were male and 36 were female.

Table 8
Role of Native Language

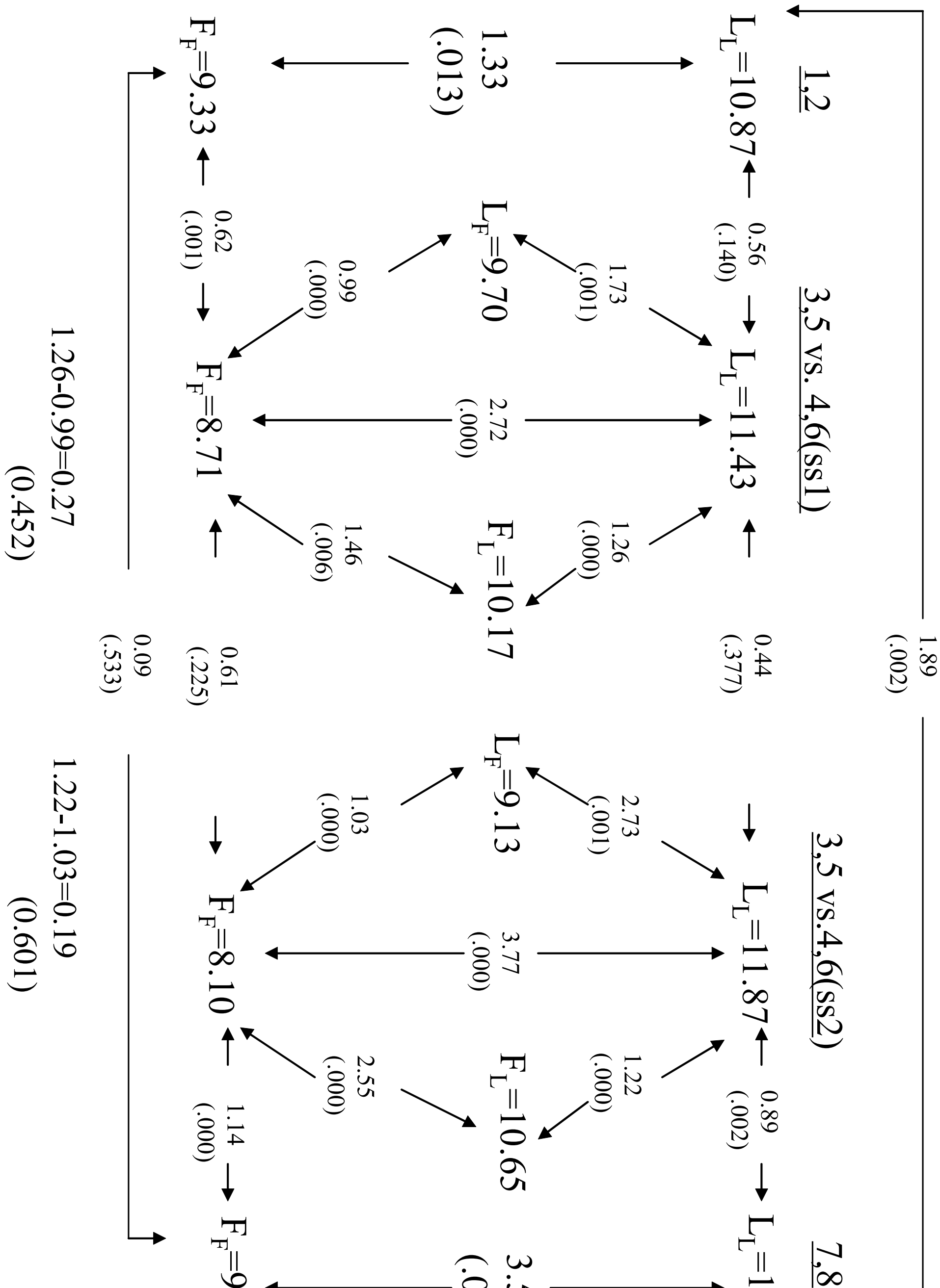
	ENGLISH (N=85)	NON-ENGLISH (N=30)	DIFFERENCE	P-VALUE
Productivity (3,4,5,6)	10.43	8.76	1.67	0.003
Linear Productivity (4,6)	11.07	9.10	1.97	0.001
Flat Productivity (3,5)	9.79	8.42	1.37	0.023
Risk Aversion*	6.88	6.87	0.01	0.978

*Of those whose responses were consistent, 76 were native English speakers and 23 were not.

Table 9
Within-Person Comparison of Productivity Under the Flat Versus the Linear Compensation Scheme: Paired Samples Statistics

ROUNDS	VANCE AND COLELLA MEANS	PRE-TEST MEANS	LINEAR	FLAT	DIFFERENCE	P-VALUE
Linear (4&6) vs Flat (3&5)	9.24/9.37	12.05/12.12	10.56	9.43	1.13	.000
Linear (4) vs Flat (5)	8.84/9.32	11.89/12.09	10.46	9.86	0.60	.042

Figure 1.
Productivity under Different Compensation Schemes



Appendix 1: Instructions

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 create words using a list of 7 letters. There will be 9 rounds in which you will be given a list of 7 letters to create words. You will have a *Workbook* that will contain all of your work. To ensure anonymity, just write down your participant number on the cover of the *Workbook*. Please do not write your name on any of these materials. Your earnings in this experiment will depend on your performance and/or the specific compensation method applied to each round.

In creating your words, we will use the following rules:

1. It is an English word.
2. It is two or more letters long.
3. It is not a proper noun (e.g., words cannot be names or places).
4. It is made by using each of the 7 letters only once per word (e.g., if the list of 7 letters contains only one ‘g’, you cannot spell “egg”).
5. It is used in only one form (e.g., you cannot use singular and plural versions of the same word).

Here is an example. Say there are 7 letters: SADFTIB. Some examples of permissible words include: “daft”, “fit”, “fad”, “bit”, and “it”. However, the word “dad” is not permissible because the letter “d” is used twice in that word. You may use either “bit” or “bits”, but not both. “I” is not permissible because it contains fewer than 2 letters. If you are unsure if a word conforms to these rules, write it down anyway.

In each round you will create words using the 7 letters at the top of the page in the *Workbook*. You will have 3 minutes to work on each round.

Once we begin the experiment, you will not be able to look ahead to future pages or to go back to previous pages. After the last round, please fill out a short questionnaire. Please respond to the questionnaire as truthfully and as accurately as possible. The questionnaire provides the experimenter with important data. Your responses to the questionnaire are confidential and will not be revealed to anyone other than the experimenter. 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.

After you have completed the questionnaire, please raise your hand and the experimenter will escort you to another room where you will be paid your earnings.

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.

Appendix 2: Post-Experiment Questionnaire

1. What is you gender? (Please mark one circle with an X)
   Male O   Female O

2. What is your age? ______________________.

3. Where were you born? **(Specify one response only, according to present boundaries)**
   Born in Australia (Specify province or territory): ___________________
   Born outside Australia (Specify country): ___________________

4. Is English your first language? (Check one.)   O Yes
   O No, my first language is ________________

5. Please indicate your preference between Option A and Option B for the following 10 paired lottery choices. ***ONE*** of these 10 choices will be selected at random after everyone has completed and handed in the questionnaire to determine your earnings.

Option A	**Option B**	**MY CHOICE (Please Write down A or B)**
10% chance of $4.40, 90% chance of $3.52	10% chance of $8.47, 90% chance of $.22	
20% chance of $4.40, 80% chance of $3.52	20% chance of $8.47, 80% chance of $.22	
30% chance of $4.40, 70% chance of $3.52	30% chance of $8.47, 70% chance of $.22	
40% chance of $4.40, 60% chance of $3.52	40% chance of $8.47, 60% chance of $.22	
50% chance of $4.40, 50% chance of $3.52	50% chance of $8.47, 50% chance of $.22	
60% chance of $4.40, 40% chance of $3.52	60% chance of $8.47, 40% chance of $.22	
70% chance of $4.40, 30% chance of $3.52	70% chance of $8.47, 30% chance of $.22	
80% chance of $4.40, 20% chance of $3.52	80% chance of $8.47, 20% chance of $.22	
90% chance of $4.40, 10% chance of $3.52	90% chance of $8.47, 10% chance of $.22	
100% chance of $4.40, 0% chance of $3.52	100% chance of $8.47, 0% chance of $.22	

7. I am willing to take high financial risks in order to realize higher average yields.

Strongly *Strongly*
Disagree *Agree*
1------------2------------3------------4------------5------------6------------7

8. I will take a job where I get paid exclusively on a commission basis.

Strongly *Strongly*
Disagree *Agree*
1------------2------------3------------4------------5------------6------------7

Thank you very much!