Attracting Creativity: The Initial and Aggregate Effects of Contract Selection on Creativity-Weighted Productivity

Steven J. Kachelmeier
Charles T. Zlatkovich Centennial Professor of Accounting
University of Texas at Austin
kach@mail.utexas.edu

and

Michael G. Williamson
Assistant Professor
University of Texas at Austin
michael.williamson@mccombs.utexas.edu

This version June 2008
Revision currently in progress

We gratefully acknowledge funding from a McCombs School of Business Research Excellence Grant and support to the first author from the Charles T. Zlatkovich Centennial Professorship. We received helpful suggestions on previous drafts from Sarah Bonner, Larry Brown, Steve Glover, Lynn Hannan, Bob Libby, Bill Messier, Mark Nelson, Tatiana Sandino, Jen Winchel, and workshop participants at Brigham Young University, Cornell University, Georgia State University, Harvard University, the University of Pennsylvania, the University of Southern California, and the Negotiation Research Center of the University of Texas at Dallas. We also wish to thank Morrie Schulman of the University of Texas at Austin Division of Instructional Innovation and Todd Pinney of Turning Technologies, Inc. for providing and tailoring the application of radio-frequency response devices for use by our creativity raters.

Attracting Creativity: The Initial and Aggregate Effects of Contract Selection on Creativity-Weighted Productivity

Abstract: Using an experiment in which participants design "rebus puzzles," we extend recent research on creativity-weighted productivity (i.e., quantity weighted by creativity ratings) by allowing participants to choose between a contract that rewards creativity-weighted productivity or one that rewards quantity only. As such, we examine both of the factors that agency theory suggests can arise from contingent compensation: (1) influencing effort (to address the moral-hazard problem of hidden action) and (2) attracting ability (to address the adverse-selection problem of hidden information). We find that participants who choose a creativity-weighted pay scheme have greater self-perceived creativity than those who choose a quantity-only scheme, and that this perceived creativity advantage manifests itself in significantly higher creativity-weighted productivity scores in initial production. For production as a whole, however, we observe a different pattern. Namely, whether compensation contracts are randomly assigned or self-selected, participants operating under a quantity-only scheme eventually produce just as many high-creativity puzzles as their creativity-weighted counterparts, and also produce significantly more puzzles overall. Thus, the implications of contract selection on creativity-weighted productivity hinge on the importance of the "head start" attained by participants who self-select a creativity-weighted contract.

Keywords: Creativity, incentives, contingent compensation, multi-dimensional performance measurement, adverse selection.

Data Availability: Please contact the authors.

Attracting Creativity: The Initial and Aggregate Effects of Contract Selection on Creativity-Weighted Productivity

I. INTRODUCTION

Success in many firms derives from efficient production and from creative efforts that enhance the value of that production (Chang and Birkett 2004; Fallon and Senn 2006). To the extent that firms desire both volume and creativity, a natural accounting solution is to incorporate both measures in a multi-dimensional performance metric (e.g., Kaplan and Norton 1996). Yet, Ittner et al. (1997) report that very few firms include explicit measures of creativity or innovation in performance-based compensation schemes, raising the question of whether there is a mismatch between what firms value and what they measure, and motivating the importance of studying the effects of creativity-weighted compensation on creativity-weighted productivity.

Recent research by Kachelmeier, Reichert, and Williamson (2008) (henceforth KRW) addresses part of this challenge. In an experimental task in which participants design "rebus puzzles," KRW find that even if creativity can be measured reliably by independent raters, creativity-weighted compensation does not maximize creativity-weighted productivity. Rather, participants whose compensation depends only on quantity attain the highest creativity-weighted productivity scores, significantly outproducing participants assigned to a creativity-weighted pay scheme in terms of overall volume, while also producing just as many high-creativity puzzles.

While their results are intriguing, KRW's experiment does not address the potential for creativity-weighted compensation to *attract* creative workers. The current study addresses this limitation by manipulating whether or not participants can choose their desired compensation scheme. From a management accounting perspective, contract selection is important because the decision-influencing value of performance-contingent compensation arises from two sources:

(1) mitigating the moral-hazard problem of hidden effort, and (2) mitigating the adverse-selection problem of hidden information (Demski and Feltham 1978; Sprinkle and Williamson 2007, §2). Even if a task depends entirely on ability rather than effort, performance-based compensation can still benefit the firm by screening prospective employees, inducing truthful revelation of abilities via contract selection.

Prior experiments have demonstrated incremental benefits from contract selection (e.g., Chow 1983; Waller and Chow 1985; Shields and Waller 1988; Dillard and Fisher 1990; Cadsby et al. 2007). However, those experiments involve alternative ways to reward a single performance dimension -- typically the quantity produced. To our knowledge, the current experiment is the first to examine contract selection in a *multi-dimensional* setting, comparing a contract that rewards only quantity to one that rewards both quantity and creativity.

We examine the effects of contract selection on creativity-weighted productivity by adapting KRW's "rebus puzzle" task to an experimental setting in which participants are either randomly assigned to contracts or are allowed to choose. For simplicity and clarity in interpreting selection preferences, we restrict attention to two of the incentive contracts considered by KRW: a piece-rate contract that rewards only quantity, or a piece-rate contract that rewards quantity weighted by creativity ratings (i.e., creativity-weighted productivity).

We detect three primary findings. First, in the random-assignment condition, we replicate KRW's primary result that participants assigned to quantity-only compensation generate significantly higher creativity-weighted productivity scores than do those who are paid for creativity-weighted productivity. This finding is noteworthy because unlike KRW, we inform participants of both potential contracts before a random device assigns them to one of these contracts, and we also provide a practice period before assigning contracts. Thus, our random-assignment results extend the KRW findings to a setting with practice opportunities and explicit knowledge of multiple possible incentive schemes.

Turning to the contract-selection condition, our second primary finding is that participants who choose the creativity-weighted pay scheme indicate significantly higher self-perceptions of creative ability and generate significantly higher creativity-weighted productivity scores than those who choose a quantity-only contract across the first several units of production. The lack of any comparable "head start" effect of creativity-weighted compensation in the random-assignment condition allows us to attribute this finding to the incremental influence of contract selection.

Third, notwithstanding the "head start" enjoyed by participants who choose the creativity-weighted pay scheme, those who choose a quantity-only scheme eventually catch up and pass. That is, creativity-weighted participants are unable to sustain their initial advantage, as quantity-only participants eventually produce just as many high-creativity puzzles and significantly more puzzles overall. Thus, for production as a whole, we see the same eventual advantage of quantity-only compensation in the contract-selection condition as in the random-assignment condition.

While the eventual dominance of quantity-only pay is consistent with the results from KRW's more basic setting, our finding of a "head start" effect of self-selected creativity-weighted compensation suggests important incremental insights beyond their study. Real-world environments often confer important advantages to those who enjoy initial success, while hindering those whose initial efforts are less impressive. Thus, the eventual ability of quantity-motivated participants to catch up with and pass their self-selected creativity-weighted counterparts might not be possible in environments that promote early success and penalize those whose creativity takes longer to develop. The ability to draw such inferences from a 20-minute production exercise is of course limited, but our results nevertheless suggest the importance of considering both initial and eventual productivity when evaluating the effectiveness of creativity-weighted compensation.

We develop theory-based hypotheses in Section II, explain our method and design in Section III, present results in Section IV, and conclude in Section V.

II. BACKGROUND, THEORY, AND HYPOTHESES

Background: Creativity-Weighted Productivity

KRW's recent investigation of the effects of performance-contingent compensation on creativity-weighted productivity provides a task and setting that we can adapt to our incremental consideration of endogenous contract selection. In their experiment, participants design "rebus puzzles," using words, diagrams, and/or pictures to depict familiar terms or phrases. This task affords meaningful measures of quantity and creativity, with the latter dimension defined in the instructions as "puzzles that are original ideas, innovative and clever," as evaluated by an independent group of raters. We describe the task in more detail in Section III. Rebus puzzles serve as an experimental analog to the "production of ideas," characterizing settings in which both quantity and creativity are important elements of productive output. Prototypical examples of such settings include professional service firms (Chang and Birkett 2004) and universities (Dewett and Denisi 2004), though KRW argue that the importance of creativity-weighted productivity is likely to generalize to many other organizational environments.

Consistent with KRW, we define "creativity-weighted productivity" as the sum of the creativity scores awarded to all puzzles produced by a participant in the allotted time. This measure is strictly increasing in both quantity and creativity, and is arithmetically equivalent to the product of quantity and the average creativity rating of that quantity. Below we consider two effects that performance-contingent compensation could have on creativity-weighted productivity. First, contingent compensation could exert "effort-influencing" effects by motivating workers to work harder at quantity, creativity, or both. This is the primary question addressed by KRW. Second, contingent compensation could exert "contract-selection" effects by attracting workers whose abilities align best with the compensated measure. This possibility is our incremental contribution.

Effort-Influencing Effects of Creativity-Weighted Compensation

KRW compare creativity-weighted productivity scores across four randomly assigned contracts: (1) a fixed-payment contract used as a control condition, (2) a contract that rewards average creativity ratings (but not quantity), (3) a contract that rewards quantity without any weighting for creativity, and (4) a contract that rewards creativity-weighted productivity. They find that the highest creativity-weighted productivity scores are generated by participants with quantity-only incentives. Participants with creativity-weighted incentives score significantly lower, producing fewer puzzles overall without generating any compensating gains in high-creativity puzzles, as proxied by puzzles rated in the overall top quartile. Participants with fixed or creativity-only incentives score lower still.

To facilitate our subsequent consideration of contract-selection effects, we restrict attention to the quantity-only and creativity-weighted productivity contracts examined by KRW. Their conclusion that creativity-weighted compensation does not maximize creativity-weighted productivity is paradoxical, because it runs counter to the conventional wisdom that people maximize what is measured (Kerr 1975). In KRW, participants paid for creativity-weighted productivity could have significantly increased the measure for which they were compensated by emulating the actions of those paid for quantity only.

KRW develop a two-step theory to explain this result. First, they draw on the judgment and decision-making literature to observe that people tend to simplify difficult tradeoffs by employing a more noncompensatory strategy, prioritizing on a condition that must be met before considering other goals (Payne et al. 1993, Ch. 2). Instead of maximizing creativity-weighted productivity, a simpler, more noncompensatory strategy when faced with this incentive is to focus efforts on high-creativity ideas, thereby overlooking the potential gains from less creative production when high-creativity ideas are unavailable.[1]

Even if a noncompensatory strategy of prioritizing on high-creativity output threatens *overall* productivity, participants employing such a strategy would still capture benefits relative to those pursuing a quantity-only strategy if prioritizing on creativity generates a greater number of high-creativity ideas. This observation is where the second part of KRW's theoretical structure applies. They draw on psychological research on creativity in noting that unlike other dimensions of performance, one cannot necessarily become more creative simply by trying harder (Amabile 1996). Thus, the "shotgun" strategy employed by quantity-only participants to maximize the production of rebus puzzles could result in as many high-creativity puzzles as the "rifle" strategy employed by creativity-weighted participants to focus on high-creativity ideas.

As a baseline condition, we reexamine these findings by randomly assigning participants to a compensation contract that rewards quantity only or one that rewards creativity-weighted productivity. Although we use the same basic task as KRW, our baseline setting with random contract assignment is not identical to their study for two primary reasons. First, in order to

[1] Creativity serves as a natural prioritizing criterion in this task (i.e., first restrict efforts to creative ideas, and then do as many as possible). Conversely, "quantity" is not a characteristic of a puzzle, and therefore cannot logically serve as a qualifying criterion (i.e., one cannot first do as many puzzles as possible, and then make them creative).

maintain *ceteris paribus* comparability to the contract-selection condition that we consider next, we describe *both* contracts to all participants before a random device assigns them to one of these contracts. In KRW, participants knew only of the compensation contract to which they were assigned. Awareness of both contracts could sensitize participants to the fact that one contract rewards creativity at the margin, whereas the other does not. Second, unlike KRW, we provide a practice period before assigning contracts. The intent of this practice period is to provide a more informed basis for contract selection, discussed shortly, such that a *ceteris paribus* comparison compels us to provide the same practice opportunity in the random-assignment condition. Due to these differences in administration, it is not self-evident that random contract assignment in our setting will replicate KRW's result, so we test that result as our first predicted hypothesis:[2]

> H1: In a setting with randomly assigned compensation contracts, participants compensated for creativity-weighted productivity will generate lower creativity-weighted productivity scores than will participants compensated only for quantity.

Contract-Selection Effects of Creativity-Weighted Compensation

Random assignment allowed KRW to isolate the effort-influencing effects of creativity-weighted compensation, controlling for self-selection biases that might otherwise contaminate a nonrandomized design. Indeed, random assignment to treatment conditions is often hailed as the primary internal validity strength of experimentation (Shadish et al. 2002). Yet, for incentive contracting, self-selection effects are not just a "confound," but rather constitute a separate reason for using such contracts that is just as important to consider as the motivational effects of contracts on effort. Namely, if workers have private knowledge of their abilities, contingent contracting can induce credible revelation of that information if different contracts appeal to workers of different abilities (Demski and Feltham 1978; Sprinkle and Williamson 2007, §2.1.1).

[2] KRW also discuss agency theoretic counterarguments against the prediction in H1, leading them to motivate the effect of incentive contracts on creativity-weighted productivity as a two-tailed research question. Given KRW's findings, however, we state H1 as an *ex ante* directional prediction for the current study.

This contract-selection benefit holds even for tasks that are insensitive to effort, so long as prospective workers have self-insight into some ability-based dimension of task performance.

Waller and Chow (1985) model the self-selection benefits of performance-contingent compensation schemes as a function of prospective employees' perceptions of their abilities and preferences. Consistent with the premise that individuals have reliable self-insights, several prior experiments have detected incremental benefits of self-selected contracts over the effort-influencing effects of such contracts (Chow 1983; Waller and Chow 1985; Shields and Waller 1988; Dillard and Fisher 1990; Cadsby et al. 2007). While these studies examine different tasks and different elements of performance-based compensation, a common theme is that they focus on a single dimension of production – the quantity produced. Thus, it remains an open question whether contract-selection benefits would extend to a multi-dimensional task, particularly when the dimension beyond quantity is subjectively determined, as is the inherent nature of creativity.

Given the complexity of a multi-dimensional setting, we present participants with only two contracts. One contract rewards quantity, whereas the other rewards creativity-weighted productivity. Importantly, the choice is not simply between quantity and creativity, as both contracts depend on quantity. That is, an individual perceiving himself/herself to have low quantity skills would not necessarily be better off with the creativity-weighted contract, because if that individual had low creative ability as well, the low quantity score would be compounded by a low creativity score. Thus, given these contracts, self-perceived creativity should be the primary factor that influences choices at the margin. Our second hypothesis tests whether individuals with higher self-perceptions of creative ability are more likely to prefer the creativity-weighted contract than the quantity-only contract:

H2: In the contract-selection condition, participants who select the creativity-weighted compensation contract perceive themselves to have greater potential for producing creative output than do individuals selecting quantity-only incentives.

If H2 holds, it is conceivable that the benefit of contract selection could offset or reverse the KRW finding that a quantity-only contract generates higher creativity-weighed productivity scores than does a contract that rewards creativity-weighted productivity. For contract selection to challenge this finding, two things must happen. First, participants' *perceived* creativity must align with their *actual* creativity. Second, contract-selection benefits must dominate the effort-influencing harm of creativity-weighted compensation that is predicted in H1. We discuss both of these possibilities in the paragraphs below.

Prior research suggests that individuals have reasonable self-insight into their creativity (Redmond et al. 1993), although that evidence is from a single creative task, and hence may not generalize to a setting in which multiple creative ideas are necessary for success. In general, Vaughn (1999) finds that individuals' self-perceptions of ability depend on how easily they can think of successful ideas, consistent with research on the "availability heuristic" (Tversky and Kahneman 1973; Kahneman and Frederick 2002; Schwarz and Vaughn 2002). Extended to our setting, Vaughn's (1999) reasoning suggests that participants' perceptions of their potential to produce creative rebus puzzles are likely to reflect their ability to think of a few creative ideas.

To the extent that individuals use the availability heuristic to assess their creative potential, we would expect those selecting the creativity-weighted contract to produce more creative output initially, reflecting the creative ideas that led them to select this contract. Whether such individuals could sustain any such initial creativity advantage is another question, because the effort-influencing effects of creativity-weighted compensation likely remain relevant under self-selected contracts. *Ex ante*, it is unclear whether any effort-influencing harm of creativity-weighted compensation (H1) will offset or dominate any benefit from allowing those with relatively high self-perceived creativity to choose the creativity-weighted pay scheme (H2). Accordingly, we state H3 below as a two-tailed hypothesis of a difference, which could arise in either direction depending on which effect is more pronounced.

> H3: In a setting in which participants can choose their desired compensation contract, participants compensated for creativity-weighted productivity will generate different creativity-weighted productivity scores than will participants compensated only for quantity.

III. METHOD

Participants and Experimental Task

In each of 10 sessions, 8-10 volunteers participated in a 75-minute experiment, yielding a total of 90 participants. We recruited participants similar to those used in KRW (i.e., students in undergraduate core business classes, reflecting a variety of business majors) and adapted KRW's basic task, subject to some modifications that we describe shortly. The instructions to both studies define a "rebus puzzle" as a "kind of riddle in which words and/or diagrams are used to represent a familiar term or phrase," offering the same 11 examples as illustrations (see Appendix). Also common to both studies is the instruction that "while we do not place any rules on the kinds of rebus puzzles you can submit, we value both the number of different puzzles you construct (i.e., quantity) and the creativity of those puzzles (i.e., puzzles that are original ideas, innovative, and clever)." This instruction is important because our interest is in the incremental effects of performance-based compensation, not the more basic goal-setting effects from knowing what matters to the experimenter (Shalley 1991; Bonner and Sprinkle 2002).

Participants worked independently at separate tables, designing puzzles on 3 × 5 inch index cards by placing the puzzle on one side and its solution on the other side. Participants placed each completed puzzle in an individual output box, subject to the understanding that once in the box, a puzzle could not be removed.[3] This rule maintained control over the sequence of submissions, allowing us to analyze creativity-weighted productivity scores over time.

Design

In a 2 × 2 between-subjects design, we manipulate (1) whether the basis of performance-contingent compensation is quantity only or creativity-weighted productivity, and (2) whether contracts are randomly assigned or chosen by the participants. Participants in all 10 sessions first read the same information about *both* incentive contracts, as reproduced in the Appendix. This procedure is an important departure from KRW, who only informed participants about the randomly assigned contract particular to their session. While our procedure retains the essence of a between-subjects design, as explained below, the common knowledge of both treatment conditions also captures the more salient comparative reference of a within-subjects design (Schepanski et al. 1992), thereby holding constant any behavioral influence of that knowledge across the random-assignment and contract-selection conditions.

We operationalize the contracts in a manner similar to KRW. The instructions (see Appendix) specify that those using the quantity-only contract will receive compensation based on a rate per puzzle that results in $5.00 for the participant (among all participants using this contract) submitting the fewest puzzles and $45.00 for the participant submitting the most puzzles, thereby yielding an average compensation of approximately $25.00. The alternative creativity-weighted contract specifies compensation based on the "creativity-weighted total score of all rebus puzzles" constructed, determined by adding the 1-to-10 creativity ratings assigned to those puzzles by "a group of students from a different class." The instructions in this version

[3] Participants were allowed to hold on to a puzzle as long as they wished before submitting it to the output box, but substantially all participants submitted puzzles to the box sequentially as they finished them, with the few instances of discarded output reflecting incomplete efforts. This behavior is reasonable, given that each completed puzzle strictly benefits compensation under either a quantity-only or creativity-weighted productivity pay scheme.

observe that "each puzzle you submit helps your total score, but higher rated puzzles count more." Similar to the quantity-only payment scheme, the creativity-weighted scheme anchors total compensation at $5.00 and $45.00 for the lowest and highest creativity-weighted total scores, respectively, yielding an average payment around $25.00.

Anchoring the low and high payments for both contracts at $5 and $45, respectively, allows us to vary the *basis* of compensation while holding constant the expected average *magnitude* of compensation. Similar to KRW, we achieve this equivalence by basing the low and high payments on the lowest and highest performers within each contract. However, we would not classify these contracts as "tournament" schemes in the sense used by Bonner et al. (2000), because with the sole exception of the highest-scoring participant, every other participant can earn strictly greater compensation by increasing the compensated performance measure, irrespective of ordinal ranking.

Our second manipulation is whether participants are randomly assigned to one of the contracts or are allowed to choose, with five of the ten experimental sessions conducted within each condition in a counterbalanced order. In the random-assignment condition, participants draw a wooden ball that is marked "A," representing the quantity-only contract, or "B," representing the creativity-weighted contract. As noted in the instructions, an equal number of balls of each type ensures that the probability of being assigned to either contract is 50 percent. In the contract-selection condition, each participant reads the same information about both contracts and writes a checkmark next to the preferred contract when ready. While contract selection does not provide any assurance of similar sample sizes, our participants were nearly evenly split, with 24 (21) of the 45 contract-selection participants choosing the quantity-only (creativity-weighted) contract, yielding relative percentages of 53 and 47 percent, respectively.

Procedure

After describing the task and the statement about valuing both quantity and creativity, but *before* describing the compensation schemes, the instructions announced a five-minute practice period "to help you gauge your quantity and creativity in constructing rebus puzzles." The intent of the practice period is to give contract-selection participants a more informed basis for choosing among contracts, so we provide the same opportunity in the random-assignment condition for *ceteris paribus* comparability. The opportunity to practice is another departure from KRW, which is why we replicate their test of randomly assigned contracts before considering the incremental effect of contract selection.

After the five-minute practice period, the experimenters collected all rebus puzzles constructed during this period and distributed new instructions about the compensated phase of the experiment. These new instructions informed participants that they would now "design rebus puzzles for an additional 20 minutes for money." To avoid ambiguity about duplicating ideas from the practice phase, the instructions informed participants that "you may reuse any of your ideas from the previous five-minute phase if you choose to do so, but you must prepare a new card for each such idea." Participants then read descriptions of both compensation schemes and were either randomly assigned to one of these contracts or indicated the preferred contract, as described above. Next, participants completed a pre-experimental questionnaire to confirm their understanding of the pay schemes and to elicit self-perceived abilities (used to test H2).

The 20-minute compensated production phase of the experiment followed. Participants then completed a post-experimental questionnaire, followed by a risk-preference exercise adapted from Boylan and Sprinkle (2001). The risk-preference instrument required participants to make 15 choices between a certain payment of $5.00 or participating in a lottery that paid $10 with probability *p* and $0 with probability 1-*p*, with *p* decreasing in five-percent increments from 85 percent to 15 percent. As a measure of risk preferences, we recorded the point at which each participant's choice shifted from preferring the lottery to preferring the certain payment of $5.00 (participants were allowed to shift only once). Earlier shifts are recorded as lower scores on this measure, reflecting lower tolerance for risk. After participants completed the questionnaire, we played out the lottery for one of the fifteen choices, selected at random. We then added either the lottery outcome or the certain $5.00 to each participant's compensation, depending on that participant's choice.

Approximately two weeks after the experimental sessions, we paid participants privately in accordance with their task performance, corresponding compensation scheme, and any additional earnings from the risk-preference instrument. As announced in the instructions, the delay in payments was necessitated by the need to determine quantity counts and creativity ratings as the basis for compensation. Payments were determined as described in the instructions, without deception of any form.

Determining Creativity Ratings

Undergraduate student volunteers (n = 40) from a different class rated the creativity of the rebus puzzles.[4] In each of four three-hour rating sessions, 10 different raters evaluated a subset of the 2,178 puzzles produced by the experimental participants (324 puzzles during the practice phase and 1,854 during the compensated phase). Raters were blind to the treatment conditions and identities of the participants, and were paid $50 for their time.

[4] Our use of undergraduate raters differs from KRW's use of doctoral students. We felt that undergraduate peer ratings would be just as reliable as doctoral-student ratings. The use of undergraduates allowed us to recruit a larger number of raters, given the need to evaluate approximately 2,000 puzzles for creativity. We have no reason to believe that the undergraduate ratings from our study differ systematically from the doctoral-student ratings in KRW, but the more important point is that we hold the rater pool constant across both the random-assignment and contract-selection conditions, such that if we replicate KRW's primary result in the random-assignment condition, we gain assurance that our contract-selection results are not simply an artifact of different creativity raters.

To construct the subset of rebus puzzles evaluated by each cohort of raters, we first removed all practice-phase puzzles that were reused by the participants during the compensated phase of the experiment. This procedure ensured that participants would not be penalized for reusing an idea from the practice period, as was explicitly allowed in the instructions. For analysis purposes, we assign the same rating to each removed practice puzzle as the rating of the corresponding compensated puzzle that reused the idea. This process reduced the total number of evaluated rebus puzzles from 2,178 to 1,968.

We then randomized the order of the 1,968 puzzles. To reduce rater fatigue and make the task manageable, each of the four rating sessions processed one-fourth of the randomly ordered puzzles. We have no reason to suspect any systematic differences between rating sessions, and even if such differences did exist, they would likely not interact with the treatment manipulations represented within each rating session. Nevertheless, to gain comfort of similar reliability across rating sessions and to give each session the same starting point for calibration, we began each rating session with the same 60 puzzles, in the same order. Correlations between the average rating within each rating session and the average across the other three rating sessions are significantly positive for the first 60 puzzles, ranging from 0.79 to 0.89. We therefore believe that the creativity ratings are reasonably reliable and consistent across sessions.[5]

Each rating session took place in a classroom, in which raters used radio-frequency response devices to assign a 1 (lowest) to 10 (highest) creativity rating to each puzzle, projected one at a time using a digital document camera. An assistant transcribed the solution to each puzzle on the bottom of each card beforehand, such that the puzzle solutions were projected along with the puzzles. The response-device software informed us when all ratings had been

[5] For analysis purposes, we obtain a creativity rating for the first 60 puzzles rated in all four sessions by randomly assigning 15 of these puzzles to each of the four sessions.

entered into a computer database, upon which one of the experimenters projected the next puzzle. To ensure independent ratings, we spaced raters far apart and asked them to refrain from comments or laughter. All raters first read the task instructions given to the experimental participants, including the same instructional examples. Raters were unaware, however, of the different compensation schemes or the assignment of participants to those schemes.[6]

Within each rating session, correlations between each rater and the average of the other nine raters are significantly positive (ranging from 0.20 to 0.73), with the exception of one rater who exhibits an insignificant negative correlation ($r = -0.06$). To enhance reliability, we drop the lowest correlated rater within each rating session. Cronbach's Alpha is 0.79 under this approach, which exceeds typical reliability thresholds (Peterson 1994).

As illustrations, Figure 1 provides examples of six rebus puzzles from our experiment, with four examples of high-creativity puzzles (Panels A through D), one example of a puzzle at the median of the creativity distribution (Panel E), and one immediately above the bottom quartile (Panel F).[7] For the entire distribution of puzzles, average ratings (nine raters per puzzle) range from 1.44 to 8.89, reflecting a wide variation in creativity.

IV. RESULTS

Overview

Before turning to our individual hypotheses, we conduct an omnibus analysis of aggregate creativity-weighted productivity. Table 1 reports descriptive statistics by condition in Panel A and a 2 × 2 ANOVA in Panel B, with contract type (quantity-only or creativity-weighted productivity) and contract determination (random assignment or self-selection) as treatment

[6] Unobtrusive identification codes on the back of each puzzle were not projected to raters.
[7] The relatively low-rated puzzle in Panel F of Figure 1, "crossing paths," follows a pattern similar to one of the instructional examples, "cross roads." KRW observe that creativity raters tend to assign lower ratings to puzzles that are patterned after one of the instructional examples.

factors.[8] The ANOVA indicates a significant main effect of contract type (quantity-only > creativity-weighted), no main effect of contract determination, and no interaction.

To explore the components of creativity-weighted productivity, Figure 2 divides the total quantity in each experimental cell into high-creativity puzzles (defined similarly to KRW as puzzles in the overall top quartile of creativity ratings, which in our distribution requires a rating of 5.7 or higher) and all other puzzles. Within both the random-assignment and contract-selection conditions, Figure 2 shows that quantity-only participants produce just as many high-creativity puzzles as do participants paid under the creativity-weighted contract, and produce significantly more puzzles overall – a pattern similar to that reported by KRW. Consistent with Figure 2, Table 2 indicates that creativity-weighted incentives significantly impair overall quantity, whereas Table 3 shows that the relative frequency of high-creativity puzzles is statistically indistinguishable across the four experimental cells. Neither table indicates any main or interactive effect of contract selection.

At first blush, these findings appear to indicate a similar dominance of quantity-only compensation in the random-assignment and contract-selection conditions. As we report shortly in more detailed tests of H2 and H3, however, that conclusion is premature. In subsequent analyses, we find that participants who select a creativity-weighted pay scheme both perceive and *achieve* a significant creativity advantage over those who select quantity-only compensation, although that advantage is restricted to the initial units of production. Thus, a time-sensitive analysis reveals different results for initial versus aggregate creativity-weighted productivity.

---

[8] Given that one of the manipulations varies whether we allow self-selection into treatment conditions, we prefer a simple ANOVA over using covariates that could influence selection choices and hence mask the effect of this manipulation. That said, our conclusions are robust to various ANCOVA approaches (not tabulated).

H1: Effort-Influencing Effects of Creativity-Weighted Compensation

H1 is our baseline replication of KRW under randomly assigned compensation contracts. To test H1, we conduct a planned comparison of creativity-weighted productivity scores between the two contract types within the random-assignment condition, finding a statistically significant difference in favor of quantity-only compensation ($t = 3.12$; one-tailed $p < .01$). Support for H1 confirms that we find the same effort-influencing effects of these contracts as reported by KRW when contracts are randomly assigned, even though we depart from KRW by describing both contracts to participants and by allowing a practice period. Following the theoretical rationale in KRW, it appears that relative to those assigned to a quantity-only contract, participants assigned to creativity-weighted incentives were less willing to work on less-creative puzzles, but were unable to generate gains in high-creativity output to compensate for their lower overall productivity.[9]

H2: Basis of Contract Selection

H2 predicts that individuals selecting the creativity-weighted compensation scheme perceive themselves to have a greater potential for producing creative output than individuals selecting the quantity-only scheme. Data for H2 are from a questionnaire administered after the practice session but before the compensated phase of the experiment, eliciting participants' self-assessments of (1) ability relative to others to produce a large quantity of rebus puzzles, (2) ability relative to others to produce highly creative rebus puzzles, and (3) quantity potential relative to creativity potential. All three questions utilize a nine-point Likert scale, with the low and high endpoints on the first two questions indicating "much worse than others" and "much

[9] Figure 2 and Table 3 indicate that if anything, quantity-only participants produced slightly *more* high-creativity puzzles than their creativity-weighted counterparts within the random-assignment condition (5.95 vs. 4.69), but this difference is not statistically significant ($t = 1.37$; two-tailed $p = .17$).

better than others," respectively. The endpoints for the third question on quantity ability relative to creative ability are 1 = "much better at quantity" and 9 = "much better at creativity."

Univariate comparisons of means reported in Panel A of Table 4 reveal that participants choosing the creativity-weighted contract perceive themselves to be significantly (1) *less* able to produce a large quantity of puzzles, (2) *more* able to produce highly creative puzzles, and (3) better at creativity than quantity. While both contracts reward quantity, such that we did not necessarily expect a difference in favor of the quantity-only contract from the quantity question, we do see the expected difference in favor of the creativity-weighted contract for the questions involving creative ability. The fact that we observe nearly identical total quantities for the quantity-only contract within the random-assignment and contract-selection conditions (25.74 vs. 25.13 – see Table 2) suggests that selection preferences have little effect on the total quantity produced by those selecting the quantity-only contract.

To test H2, Panel B of Table 4 reports a logistic regression on the propensity to choose the creativity-weighted contract, regressed on responses to the question involving creative ability and also on the supplemental measure from the exercise we used to elicit risk preferences (descriptive statistics from which are reported in Table 4, Panel A). We control for risk preferences because the subjectivity of creativity ratings implies greater uncertainty under the creativity-weighted pay scheme than under the quantity-only scheme. In general, individuals with greater tolerance for risk are more likely to prefer noisier compensation schemes (Cadsby et al. 2007), supporting the inclusion of a risk-preference measure in the logistic regression.

Logit results indicate that the propensity of participants to select the creativity-weighted contract is positively associated with self-perceived creative potential and with risk tolerance.[10]

[10] A logistic regression using self-assessments of creative potential relative to quantity potential generates inferentially identical results. We also find that selection results are insensitive to participant gender.

We also find a significant interaction between the creative potential and risk-preference measures, indicating that the combined effect of these two factors is subadditive. That is, a greater score on either one of these measures dampens (but does not eliminate) the effect of the other measure on the propensity to select the creativity-weighted contract.

Although beliefs about creative potential are significantly associated with the propensity to choose the creativity-weighted pay scheme, contract selection does not appear to depend on either the quantity or creativity of puzzles generated during the five-minute practice period. Panel A of Table 4 shows that means for quantity, average creativity ratings, and creativity-weighted productivity during the practice period do not differ significantly between contracts ($p > 0.50$ in each case). One reason why practice output might not be significantly associated with contract selection is that participants had no compensated incentives during the practice period, and did not yet know the alternative pay schemes from which they would be choosing. As we report next, however, we do discern a creativity-based advantage among those choosing the creativity-weighted pay scheme in the initial units of their *compensated* production.

H3: Effects of Contract Selection on Creativity-Weighted Productivity

We have already seen from Table 1 that participants paid under the quantity-only contract achieve higher *total* creativity-weighted productivity scores than those whose pay depends on creativity-weighted productivity. A planned pairwise comparison confirms that this difference is statistically significant within the contract-selection condition ($t = 2.97$; two-tailed $p < .01$), similar to the difference within the random-assignment condition reported earlier in our test of H1. Thus, in the aggregate, our results resolve H3 in favor of the quantity-only contract.

These aggregate results notwithstanding, a deeper consideration of the “availability heuristic” underlying H2 suggests that at the time of contract selection, those selecting the creativity-weighted contract likely do so because their initial ability to think of creative ideas exceeds that of participants who prefer the quantity-only scheme. If this reasoning is valid, we should see some corroborating evidence in our test of H3, at least in initial production.[11] Accordingly, we examine H3 in more detail by analyzing how creativity-weighted productivity develops over time. Figure 3 graphs separate plots for the random-assignment and contract-selection conditions of the *cumulative difference* in creativity-weighted productivity scores between those operating under the creativity-weighted contract and those paid for quantity only, at each unit of production. Positive (negative) values on the vertical axis indicate that the creativity-weighted contract generates higher (lower) cumulative creativity-weighted productivity than the quantity-only contract through the unit of production reflected on the horizontal axis. Thus, Figure 3 is essentially a running "horse race" of which contract is generating higher cumulative creativity-weighted productivity scores at each unit of production. For early units of production, this measure will primarily reflect relative creativity ratings.[12] As the production unit gets larger, this measure will reflect both creativity ratings and quantity, to the extent that participants fail to produce the number of puzzles indicated on the horizontal axis.

Figure 3 indicates that individuals selecting the creativity-weighted productivity contract outperform their quantity-only counterparts by a margin that increases through its peak at the ninth unit of production. Statistically, this difference is at least marginally significant for the first eight units ($t = 1.98$; two-tailed $p = .05$) and for the first nine units ($t = 1.94$; two-tailed $p = .06$), in a direction *opposite* to the aggregate creativity-weighted productivity results reported in Table 1. This initial "creative spark" reflects the incremental effect of contract selection, as

---

[11] In addition to the practice period, we gave participants three minutes to choose their desired pay scheme, affording ample opportunity to generate some initial ideas.

[12] No participant submitted fewer than six rebus puzzles, such that for horizontal axis values less than six, Figure 3 reflects the cumulative difference in average creativity ratings multiplied by the number on the horizontal axis.

Figure 3 shows a nearly monotonic tendency for quantity-only participants to outperform their creativity-weighted counterparts by an increasing degree over time within the random-assignment condition.[13] Thus, it appears that allowing participants to self-select a creativity-weighted contract does indeed attract those whose initial ideas are relatively creative.

To be sure, Figure 3 also shows that the initial advantage of the creativity-weighted contract over the quantity-only contract within the contract-selection condition eventually reverses. The two self-selected contracts generate equivalent cumulative creativity weighted-productivity scores by the 14th unit of production, and by the 16th unit and beyond, quantity-only compensation dominates creativity-weighted compensation by an increasing magnitude that is similar to that observed under randomly assigned contracts. Thus, we find that even with self-selected pay schemes, quantity-only participants eventually catch up with and pass their creativity-weighted counterparts, generating the aggregate differences in Table 1 and Figure 2.

In sum, we reach different conclusions for H3 depending on whether we focus on initial or aggregate creativity-weighted productivity. One might reason that the aggregate results are more important, but in real-world environments, a "head start" in creative endeavors can lead to important advantages. For example, extending the academic analogy suggested by KRW, universities confer tenure upon those who achieve early creative success in an academic career, while those with lesser initial success can find themselves with limited opportunities to catch up. Accordingly, we would not characterize our results as simply demonstrating the robustness of KRW to contract selection. Rather, we find that a self-selected creativity-weighted pay scheme

---

[13] More formally, the *interaction* between the contract type and contract determination manipulations is at least marginally significant for cumulative creativity-weighted productivity scores through the eighth ($F = 3.86$; two-tailed $p = .05$) and ninth ($F = 2.96$; two-tailed $p = .09$) units of production, indicating that the initial difference favoring creativity-weighted pay within the contract-selection condition statistically exceeds the difference observed in the random-assignment condition.

can indeed attract relatively creative workers who enjoy greater initial success, notwithstanding the fact that effort-influencing effects eventually favor the quantity-only contract.

Supplemental Analysis of High-Creativity Output

Because our measure of creativity-weighted productivity is strictly increasing in all output produced, it is also important to examine high-creativity output in particular. As reported previously, Table 3 analyzes high-creativity (i.e., overall top quartile) puzzles in the aggregate, finding little effect of contract selection. However, similar to the difference between initial and aggregate creativity-weighted productivity detected in our test of H3, we discern different high-creativity production over time when contracts are self-selected than when contracts are randomly assigned. Panel A of Figure 4 shows that within the random-assignment condition, participants using the quantity-only and creativity-weighted contracts generate high-creativity puzzles at a nearly identical rate across the first several units of production. Within the contract-selection condition, by contrast, Panel B of Figure 4 indicates that participants choosing the creativity-weighted contract generate high-creativity puzzles at an initially faster rate than their quantity-only counterparts. For example, through the eighth unit of production, an average of 2.52 of the self-selected creativity-weighted participants' puzzles are high-creativity efforts, as opposed to 1.75 of the puzzles generated by participants who prefer the quantity-only contract ($t = 1.89$; two-tailed $p = .06$).[14] The corresponding difference through eight units of production in the random-assignment condition is in the opposite direction if anything, though clearly not statistically significant (2.08 creativity-weighted vs. 2.21 quantity-only; $t = 0.32$; $p > .50$).[15]

---

[14] The statistical significance of the cumulative creativity-weighted vs. quantity-only difference in high-creativity production within the contract-selection condition peaks at the 11^{th} unit of production ($t = 2.22$; two-tailed $p = .03$).

[15] A 2 × 2 analysis of high-creativity output through the eighth unit of production yields a marginally significant interaction between the contract type and contract determination manipulations ($F = 2.44$; two-tailed $p = .12$), indicating that the creativity-weighted vs. quantity-only difference within the contract-selection condition exceeds the difference observed within the random-assignment condition. Including self-perceived familiarity with rebus puzzles as a covariate improves the statistical significance of this interaction to $p = .097$ (two-tailed).

Figure 4 also shows that the high-creativity production advantage of self-selected creativity-weighted participants eventually dissipates, as such participants reach their quantity limits while quantity-only participants continue to generate high-creativity puzzles along with their more modest efforts. Thus, we discern a temporal pattern for high-creativity output similar to that observed for overall creativity-weighted productivity: self-selected creativity-weighted participants enjoy an initial advantage, but quantity-only participants eventually catch up.

Highest-Creativity Output

Even the top quartile of creativity may be insufficiently selective in some environments, so as an additional supplemental test, we examine the effects of contract type and contract determination on the *highest*-rated puzzle produced by each participant. Table 5 reports results, with descriptive statistics in Panel A and a 2 × 2 ANOVA in Panel B. Panel A indicates that the average highest creativity rating by participant is in the 7.2-7.3 range for three of the four experimental cells, with the sole exception being participants who choose the quantity-only contract, for which the highest rating averages 6.91.

This difference is not dramatic, but it yields an interaction between contract type and random vs. self-selected contract determination that borders on statistical significance (two-tailed $p = .13$), with the simple effect of contract type within the contract-selection condition generating a two-tailed p-value of .10. While we do not wish to overemphasize this difference, it does provide some evidence of a sustainable advantage of self-selected creativity-weighted compensation (or disadvantage of self-selected quantity-only compensation) that emerges for production as a whole. If a participant's maximum attainable creativity is of particular importance, our findings indicate that contract selection could have a modest detrimental effect among participants who choose a contract that does not reward creativity.

V. CONCLUSIONS

In what we believe is the first experimental study to investigate contract selection in a multi-dimensional setting, we examine whether the contract-selection benefits of a pay scheme that rewards creativity-weighted productivity (i.e., total quantity weighted by creativity ratings) offset the effort-influencing costs of this scheme, relative to an alternative contract that rewards quantity only. Using the same basic "rebus puzzle" task as in a recent related study by KRW, we first replicate their finding that when contracts are randomly assigned, participants paid for quantity only generate higher creativity-weighted productivity scores than participants paid for creativity-weighted productivity. Our random-assignment condition extends theirs by describing both contracts to participants before assigning contracts and by allowing a practice period. Similar to KRW, we find that effort-influencing effects favor the quantity-only contract, insofar as participants assigned to the creativity-weighted contract produce no more high-creativity puzzles than their quantity-only counterparts, but produce significantly fewer puzzles overall.

Given the similarity of our results to KRW for randomly assigned contracts, we consider the incremental effects of allowing participants to choose between contracts. Controlling for risk-preferences, we find that participants who choose the creativity-weighted pay scheme indicate higher self-perceptions of creative ability than do those who choose the quantity-only scheme. These self-perceptions are manifest in creativity-weighted productivity scores that are initially higher among those who select the creativity-weighted contract. Eventually, however, quantity-only participants catch up with and pass their creativity-weighted counterparts, such that for production as a whole, we observe a similar effort-influencing advantage of quantity-only compensation as in the random-assignment condition.

In sum, we find that there is no simple answer to the question of whether allowing participants to select a creativity-weighted contract enhances creativity-weighted productivity. For initial production, we find contract-selection benefits consistent with the premise that a creativity-weighted contract can attract relatively creative workers. To the extent that initial creative success determines subsequent career paths and resources, this result qualifies KRW's conclusion that creativity-weighted compensation impairs creativity-weighted productivity. For aggregate production, however, we find that the effort-influencing advantages of a quantity-only scheme eventually dominate. Thus, in settings in which all workers have enough opportunity to "catch up," our results suggest that even those who self-select a quantity-only contract can eventually produce just as many high-creativity efforts as their creativity-weighted counterparts, along with several other efforts of more modest creativity. As a caveat to this finding, we also detect a marginally significant tendency for self-selected quantity-only incentives to impair the creativity rating of the highest-rated output a participant is able to produce.

Our results build on KRW's efforts to gain a better understanding of why firms are reluctant to reward creativity explicitly, despite calls to do so (Ittner et al. 1997). Certainly other elements besides contract selection could shed additional insights towards this end. For example, although we detect different results in initial versus aggregate production even in a 20-minute production exercise, truly creative efforts can often take days, weeks, or years. A long-term experiment of this genre would be more difficult to control, depending on the task, as participants might be tempted to "cheat" by accessing others' efforts in print or online, if given the opportunity. Still, our results suggest that an internally valid longitudinal analysis of creativity-weighted productivity could shed important insights. Similar comments extend to the possible moderating effects of feedback (e.g., see Sprinkle 2008), albeit limited by the inherent delays that characterize reliable creativity feedback both in the laboratory and in practice. Even without feedback, our findings suggest that participants had reasonable self-insight into their creativity, as evidenced by the initially more creative output we observe among those who select the creativity-weighted contract. We leave further exploration of these issues to future studies.

In closing, we echo KRW's call for expanding the experimental accounting literature on performance-based incentives beyond the traditional interest in quantity. Like KRW, our focus is on blending productivity with creativity, which we believe to be of particular importance in contemporary business practice. As a different example, Farrell et al. (2008) design an experiment to examine the quantity-quality tradeoff, in which "quality" captures elements of accurate alignment with customer specifications. Whether creativity, quality, or other dimensions, we believe that accounting researchers can and should incorporate "softer" elements of performance into the study of performance-based compensation. Our study evidences that laboratory methods are well-suited to such inquiries for both the moral-hazard effects of hidden actions and the adverse-selection effects of hidden information.

Appendix: Experimental Instructions

Ground rules

Before describing the experiment, it is important to establish two ground rules.

1. NO TALKING WITHIN OR BETWEEN SESSIONS

While we hope that you find this experiment to be fun, it is also serious research. Please help us maintain control over the experiment by refraining from comments or other communication with your fellow participants in this session or with other students who might be participating in future sessions. You will be working individually during this experiment, so there is no need to communicate with other participants. If you have any questions, just raise your hand and we will assist you.

2. NO DECEPTION

We promise to carry out the experiment in the manner described in these instructions, with no deception of any form. As will be explained later, we will pay your compensation for this experiment at a later date (in about two weeks), but we promise that your compensation will be determined exactly as described in the rules explained later for this session.

Task

In this research, you will be constructing "rebus puzzles." A rebus puzzle is a kind of riddle in which words and/or diagrams are used to represent a familiar term or phrase. Here are some examples:

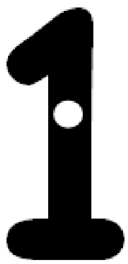
Man ——— Board	Stand ——— I	You Just Me	Just 144 ice

These examples use the positioning of the words to create the riddle. The first one is "man overboard," because "man" is over "board." The second one is "I understand," for similar reasons. The third example is "just between you and me" (i.e., "just" is between "you" and "me"), and the fourth is "gross injustice" – because the number 144 (a gross) is "in" justice.

Other rebuses use counting, different sizes, shapes, or positions to create the riddle, such as these examples:

Rebus puzzle	Solution
0 ——— M.D. Ph.D. B.A.	Three degrees below zero.
Funny Funny Words Words Words Words	Too funny for words (2 funny; 4 words)
paiNS	Growing pains
R R O A D S A D S	Cross roads
CHAIR	High chair

Still other rebuses use simple pictures, symbols, and diagrams as part of the riddle, like these two examples:

Rebus puzzle	Solution
	A hole in one
BB 	To be or not to be

What we would like you to do

At your desk is a stack of blank index cards. You will be using these cards to construct your own rebus puzzles, printing the ***solution on the back side of the card***. For instance, at your desk is a completed card with the "man overboard" puzzle from p. 1 as an example.

As you finish each puzzle, please put it in the box provided on your desk. Once you put a puzzle in the box, it is considered finished. Please do not remove any puzzles from the box after you finish them.

While we do not place any rules on the kinds of rebus puzzles you can submit, we value both the ***number of different puzzles*** you can construct (i.e., quantity) and the ***creativity*** of those puzzles (i.e., puzzles that are original ideas, innovative, and clever).

At this point we will ask you to try your hand at designing rebus puzzles for a five-minute period, trying to achieve both quantity and creativity. Please remember to write the solution to each puzzle on the back of the card.

Compensated Phase

We will now describe the compensated phase of the experiment, in which you will design rebus puzzles for an additional **20 minutes** for money, as we will explain shortly. During this phase, you may re-use any of your ideas from the previous five-minute phase if you choose to do so, but you must prepare a new card for each such idea.

While (as explained previously) we value both ***quantity*** and ***creativity*** for ***all*** participants in this research, you will be operating under one of two different compensation formulas, described as follows.

Compensation Formula A:

Under **Compensation Formula A**, your compensation will be based on ***how many*** rebus puzzles you can construct in 20 minutes. To determine this, we will count the number of rebus puzzles constructed by each person using this compensation formula.

We will then determine a cash payment per puzzle, where the payment rate per puzzle results in **$45.00** total compensation for the participant ***using this formula*** (or participants, if tied) who constructs the ***most*** rebus puzzles, and **$5.00** for the participant who constructs the ***fewest*** rebus puzzles. Everyone else using this formula will receive something in between $5.00 and $45.00, depending on the number of rebus puzzles constructed, to result in an expected average compensation around **$25.00.**

The more puzzles you submit under this formula, the more money you will make. We will pay you in about two weeks, after we have analyzed the results to determine the payment rate that achieves this compensation. We promise that you and all others using this compensation formula will receive cash compensation as described above for participating today.

Compensation Formula B:

Under **Compensation Formula B,** your compensation will be based on ***the creativity-weighted total score*** of all rebus puzzles you can construct in 20 minutes. To determine this score, we will ask a group of students from a different class to rate the creativity of each puzzle you submit on a 1 to 10 scale, where "10" is the highest possible rating and "1" is the lowest possible rating. We will ***add*** the creativity ratings on a 1 to 10 scale of ***all*** puzzles you submit. Thus, ***each puzzle you submit helps your total score under this formula, but higher rated puzzles count more*** (at the extreme, a puzzle rated 10 counts ten times as much as a puzzle rated 1).

We will then determine a cash payment rate to result in **$45.00** total compensation for the participant ***using this formula*** (or participants, if tied) with the ***highest*** total score, and **$5.00** for the participant with the ***lowest*** total score. Everyone else using this formula will receive something in between $5.00 and $45.00, depending on individual total scores, to result in an expected average compensation around **$25.00**.

The higher your creativity-weighted total score under this formula, the more money you will make. We will pay you in about two weeks, after we have analyzed the results to determine the payment rate that achieves this compensation. We promise that you and all others using this compensation formula will receive cash compensation as described above for participating today.

Summary of Compensation Formulas

Under Compensation Formula **A**, your pay will be based on ***how many puzzles you submit.*** Under this formula, ***each puzzle you submit increases your compensation***.

Under Compensation Formula **B**, your pay will be based on the ***creativity-weighted total of the puzzles you submit***, determined by adding together the creativity ratings (from 1 to 10) that we will get for all of your puzzles. Like Compensation Formula A, ***each puzzle you submit increases your compensation***, but under this formula, ***higher rated puzzles increase your pay more*** than lower rated puzzles.

Pay rates will be calculated separately for the participants under each formula, such that the ***average*** compensation for both formulas should be about $25, but will range between $5 and $45.

Which Formula Is Relevant to Me?

[Random-assignment condition]

You will determine whether you are using Compensation Formula A or B by drawing a wooden ball marked “A” or “B” from a bag. Each participant has a 50% chance of operating under either formula. With several sessions we are planning, we expect that there will be many students participating under each formula.

Please check one of the following based on your random draw:

____ I will be using **Compensation Formula A.**

____ I will be using **Compensation Formula B.**

[Contract-selection condition]

Whether you are using Compensation Formula A or B is entirely your choice. Please think about each formula and then choose either A or B by checking your preference below. With several sessions we are planning, we expect that there will be many students participating under each formula.

Please check one of the following based on your preference:

____ I will be using **Compensation Formula A.**

____ I will be using **Compensation Formula B.**

REFERENCES

Amabile, T. M. 1996. *Creativity in Context: Update to the Social Psychology of Creativity*. Boulder, CO: Westview Press.

Bonner, S. E., R. Hastie, G. B. Sprinkle, and S. M. Young. 2000. A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. *Journal of Management Accounting Research* 12: 19-64.

______, and G. B. Sprinkle. 2002. The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research. *Accounting, Organizations and Society* 27 (May – July): 303-345.

Boylan, S. J., and G. B. Sprinkle. 2001. Experimental evidence on the relation between tax rates and compliance: The effect of earned vs. endowed income. *The Journal of the American Taxation Association* 23 (Spring): 75-90.

Cadsby, B. C., F. Song, and F. Tapon. 2007. Sorting and incentive effects of pay for performance: An experimental investigation. *Academy of Management Journal* 50 (April): 387-405.

Chang, L., and B. Birkett. 2004. Managing intellectual capital in a professional service firm: Exploring the creativity-productivity paradox. *Management Accounting Research* 15 (March): 7-31.

Chow, C. W. 1983. The effects of job standard tightness and compensation scheme on performance: An exploration of linkages. *The Accounting Review* 58 (October): 667-685.

Demski, J. S., and G, A, Feltham. 1978. Economic incentives in budgetary control systems. *The Accounting Review* 53 (April): 336-359.

Dewett, T., and A. S. Denisi. 2004. Exploring scholarly reputation: It's more than just productivity. *Scientometrics* 60 (June): 249-272.

Dillard, J. F., and J. G. Fisher. 1990. Compensation schemes, skill level and task performance: An experimental examination. *Decision Sciences* 21 (Winter): 121-137.

Fallon, P. and F. Senn. 2006. *Juicing the Orange: How to Turn Creativity into a Powerful Business Advantage*. Boston, MA: Harvard Business School Press.

Farrell, A. M., K. Kadous, and K. L. Towry. 2008. Contracting on contemporaneous vs. forward-looking measures: An experimental investigation. *Contemporary Accounting Research* 25: forthcoming.

Ittner, C. D., D. F. Larcker, and M. V. Rajan. 1997. The choice of performance measures in annual bonus contracts. *The Accounting Review* 72 (April): 231-255.

Kachelmeier, S. J., B. E. Reichert, and M. G. Williamson. 2008. Measuring and motivating quantity, creativity, or both. *Journal of Accounting Research* 46 (May): 341-373.

Kahneman, D., and S. Frederick. 2002. Representativeness revisited: Attribute substitution in intuitive judgment. In *Heuristics and Biases: The Psychology of Intuitive Judgment*, edited by T. Gilovich, D. Giffin, and D. Kahneman, 49-81. New York, NY: Cambridge University Press.

Kaplan, R., and D. Norton. 1996. *The Balanced Scorecard: Translating Strategy into Action*. Boston, MA: Harvard Business School Press.

Kerr, S. 1975. On the folly of rewarding A, while hoping for B. *Academy of Management Journal* 18 (December): 769-783.

Payne, J. W., J. R. Bettman, and E. J. Johnson. 1993. *The Adaptive Decision Maker*. New York, NY: Cambridge University Press.

Peterson, R. A. 1994. A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research* 21 (September): 381-391.

Redmond, M. R., M. D. Mumford, and R. Teach. 1993. Putting creativity to work: Effects of leader behavior on subordinate creativity. *Organizational Behavior and Human Decision Processes* 55 (June): 120-151.

Schepanski, A., R.M. Tubbs, and R.A. Grimlund. 1992. Issues of concern regarding within- and between-subjects designs in behavioral accounting research. *Journal of Accounting Literature* 11: 121-150.

Schwarz, N., and L. A. Vaughn. 2002. The availability heuristic revisited: Ease of recall and content of recall as distinct sources of information. In *Heuristics and Biases: The Psychology of Intuitive Judgment*, edited by T. Gilovich, D. Giffin, and D. Kahneman, 103-119. New York, NY: Cambridge University Press.

Shadish, W. R., T. D. Cook, and D. T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.

Shalley, C. E. 1991. Effects of productivity goals, creativity goals, and personal direction on individual creativity. *Journal of Applied Psychology* 76 (April): 179-185.

Shields, M. D., and W. S. Waller. 1988. A behavioral study of accounting variables in performance-incentive contracts. *Accounting, Organizations and Society* 13 (November): 581-594.

Sprinkle, G. B., and M. G. Williamson. 2007. Experimental research in managerial accounting. In *Handbook of Management Accounting Research*, Volume 1, edited by C. S. Chapman, A. G. Hopwood, and M. D. Shields, 415-444. Oxford, UK: Elsevier, Ltd.

______. 2008. Discussion of Measuring and motivating quantity, creativity, or both. *Journal of Accounting Research* 46 (May): 375-382.

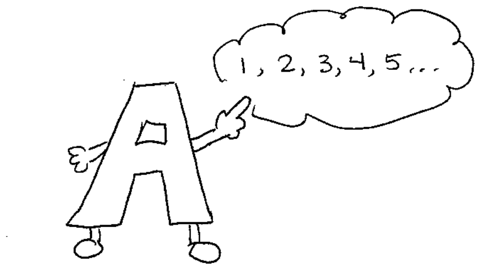
Tversky, A. and D. Kahneman. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5: 207-232.

Vaughn, L. A. 1999. Effects of uncertainty on use of the availability heuristic for self-efficacy judgments. *European Journal of Social Psychology* 29 (March): 407-410.

Waller, W. S., and C. W. Chow. 1985. The self-selection and effort effects of standard-based employment contracts: A framework and some empirical evidence. *The Accounting Review*, 60 (July): 458-476.

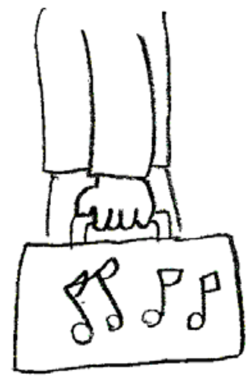
FIGURE 1
Examples of Actual Rebus Puzzles

Panel A



Solution: Accounting
Creativity rating: 8.33
Condition: Random/ Quantity

Panel B



Solution: Carry a tune
Creativity rating: 7.89
Condition: Random/ Creativity-weighted Productivity

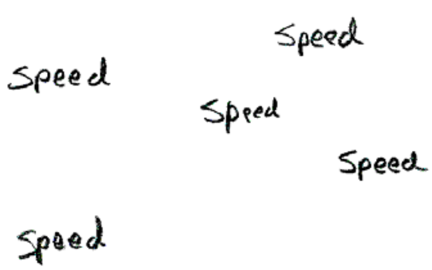
Panel C

Solution: Nail biter
Creativity rating: 7.33
Condition: Selection/ Quantity

Panel D

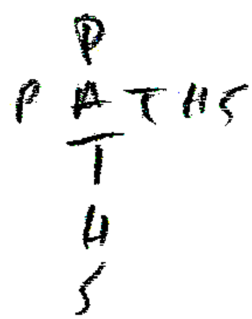
Solution: Spoiled milk
Creativity rating: 7.56
Condition: Selection/ Creativity-weighted Productivity

Panel E



Solution: Five speed
Creativity rating: 4.67 (Median Ranking)
Condition: Random/ Quantity

Panel F



Solution: Crossing paths
Creativity rating: 3.78 (25th Percentile)
Condition: Selection/ Quantity

FIGURE 2
Number of Puzzles with a Composite Creativity Rating Above and Below 5.7 by Experimental Condition

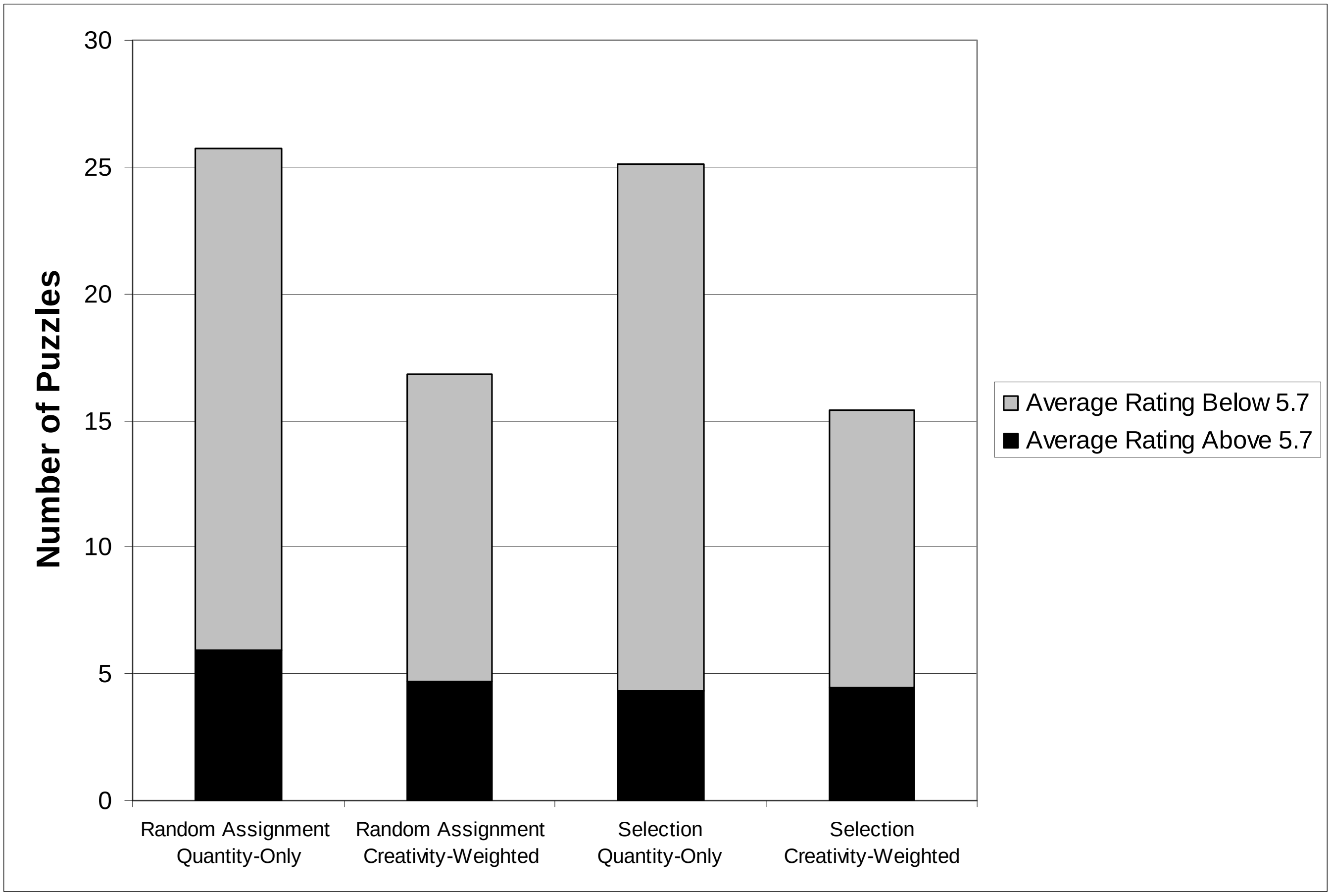
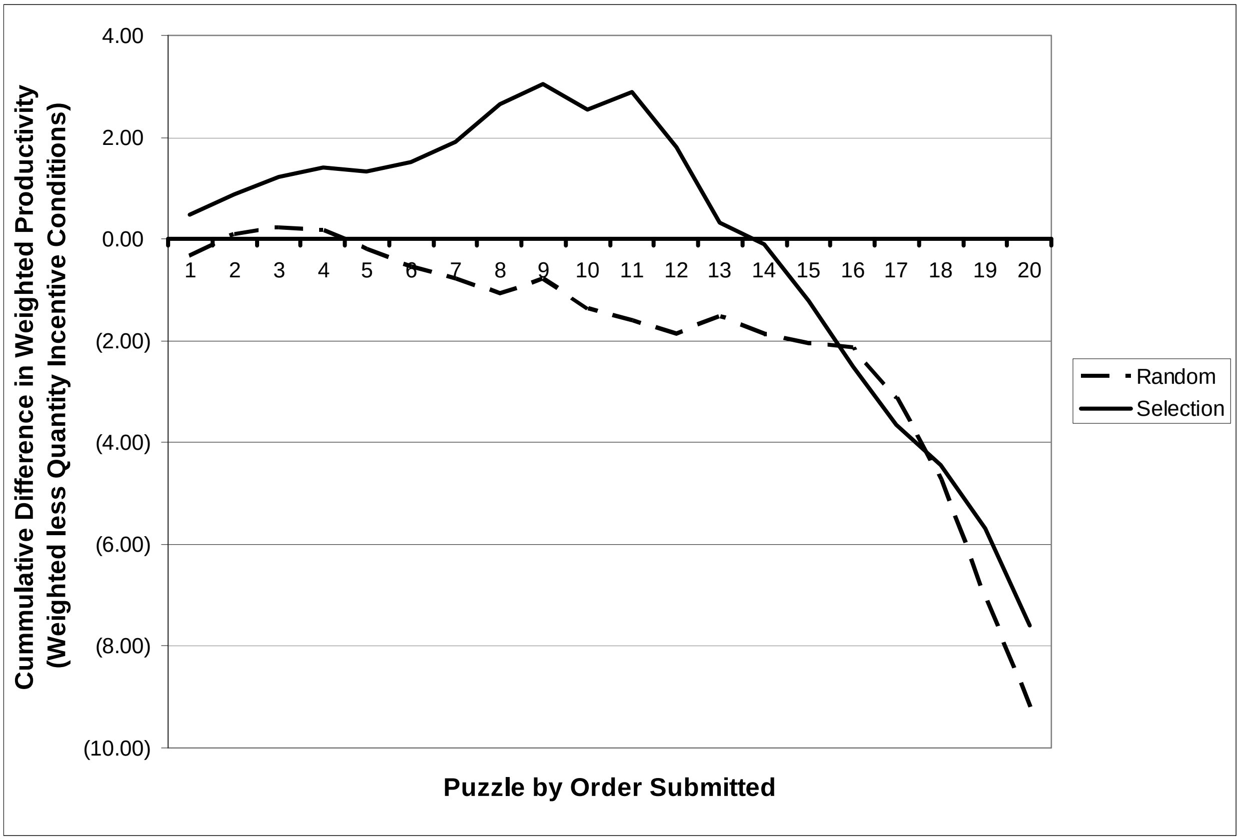


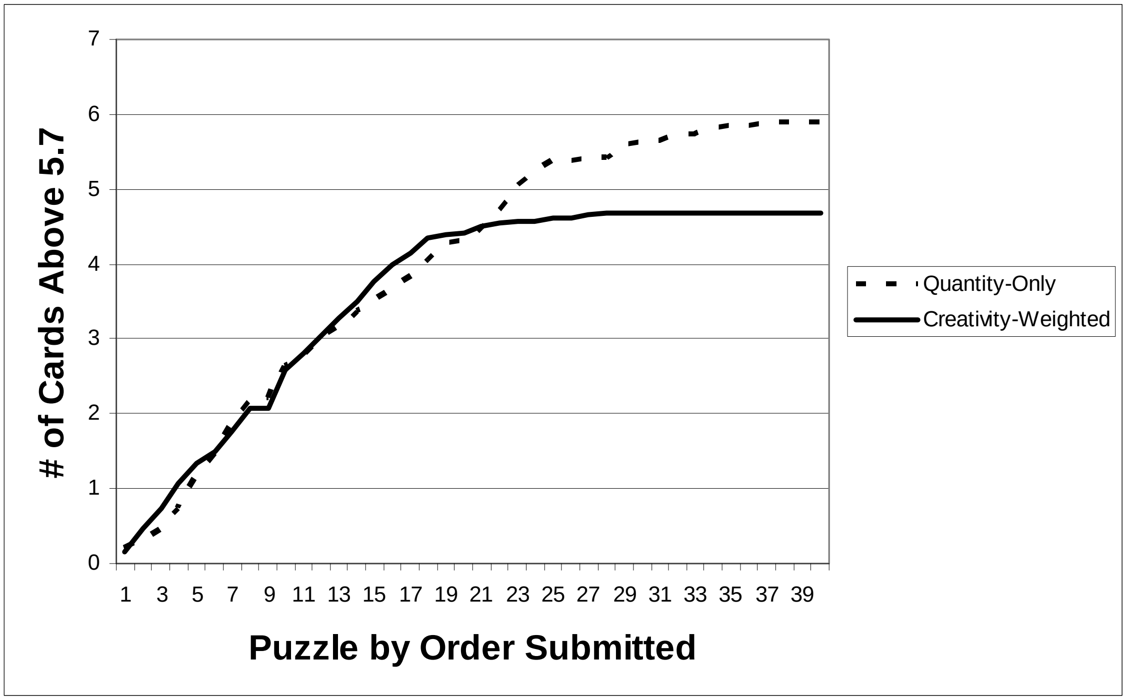
FIGURE 3
Cumulative Creativity Weighted-Productivity by Order Submitted



For both the random assignment and selection conditions, this graph presents the difference in cumulative creativity-weighted productivity between individuals with creativity-weighted productivity incentives and those with quantity incentives by order of puzzle submitted.

FIGURE 4
Cumulative Production of High-Creativity (Overall Top Quartile) Puzzles

Panel A: Random-assignment condition



Panel B: Contract-selection condition

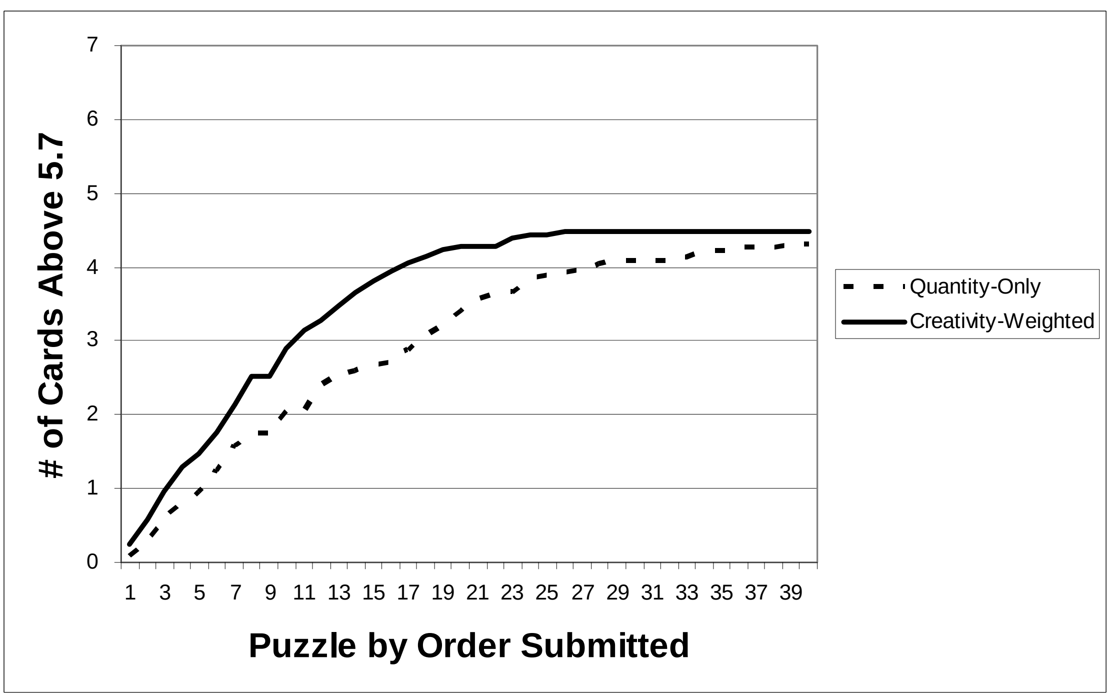


TABLE 1
ANOVA for the Effect of Contract Type and Contract Determination on Creativity-Weighted Productivity

Panel A: Means (Standard Deviation) for Creativity-Weighted Productivity

	Contract Type	
	Quantity-Only	Creativity-Weighted
Contract Determination		
Random Assignment	122.05 (40.98) n = 19	83.88 (27.07) n = 26
Selection	113.32 (57.12) n = 24	77.31 (29.92) n = 21

Panel B: Analysis of Variance

Factor	*df*	*Sum of Squares*	*F*	*p-value (two-tailed)*
CONTRACT TYPE	1	30,510.26	18.54	<.01
CONTRACT DETERMINATION	1	1,298.79	0.79	.38
CONTRACT TYPE × DETERMINATION	1	25.64	0.02	>.50
Error	86			

TABLE 2
ANOVA for the Effect of Contract Type and Contract Determination on Quantity

Panel A: Means (Standard Deviation) for Quantity

	Contract Type	
	Quantity-Only	Creativity-Weighted
Contract Determination		
Random Assignment	25.74 (10.22)	16.85 (5.42)
Selection	25.13 (15.41)	15.43 (6.06)

Panel B: Analysis of Variance

Factor	*df*	*Sum of Squares*	*F*	*p-value (two-tailed)*
CONTRACT TYPE	1	1915.31	18.70	<.01
CONTRACT DETERMINATION	1	22.83	0.22	>.50
CONTRACT TYPE × DETERMINATION	1	3.60	0.04	>.50
Error	86			

TABLE 3
ANOVA for the Effect of Contract Type and Contract Determination on Number of Puzzles Rated Above 5.7 on Creativity Scale

Panel A: Means (Standard Deviation) for Number of Cards Rated Above 5.7 on Creativity Scale

	Contract Type	
	Quantity-Only	Creativity-Weighted
Contract Determination		
Random Assignment	5.95 (4.01)	4.69 (3.06)
Selection	4.33 (2.84)	4.48 (2.06)

Panel B: Analysis of Variance

Factor	*df*	*Sum of Squares*	*F*	*p-value (two-tailed)*
CONTRACT TYPE	1	6.86	0.74	.39
CONTRACT DETERMINATION	1	18.57	2.01	.16
CONTRACT TYPE × DETERMINATION	1	10.83	1.18	.28
Error	86			

TABLE 4
Basis of Selection

Panel A: Means (Standard Deviation) for Self-assessed Potential Variables and Risk Preferences

	Contract Selected *Mean (Standard Deviation)*		**t-statistic**
	Quantity-Only n=24	Creativity-Weighted n=21	*(two-tailed p-value)*
Quantity potential[a]	5.63 (1.14)	4.19 (1.78)	3.27 (<.01)
Creativity potential[b]	3.92 (1.25)	6.14 (1.59)	5.26 (<.01)
Creativity relative to quantity potential[c]	3.67 (1.20)	6.05 (1.43)	6.06 (<.01)
Risk preference[d]	6.79 (2.43)	7.86 (3.44)	1.21 (.23)
Practice quantity	3.21 (1.44)	3.19 (1.78)	0.04 (>.50)
Practice average creativity	4.86 (0.87)	4.75 (0.76)	0.42 (>.50)
Practice weighted productivity	15.49 (7.08)	14.90 (8.18)	0.26 (>.50)

Panel B: Results of Logistic Regression Analysis for Probability of Choosing Creativity-Weighted Productivity

Independent Variable	**Estimate (Standard Error)**
Constant	-13.20 (4.75)*
Creativity potential	2.48 (0.91)*
Risk preference	0.96 (0.48)*
Creativity potential × Risk Preference	-0.17 (0.09)*

The dependent variable takes on a value of "1" for individuals choosing the creativity-weighted productivity scheme and a "0" for individuals choosing quantity-only scheme. The numbers in the table are coefficients (t-statistics). Results of two-tailed tests to determine whether the t-statistic is significantly different from zero: * = $p < 0.05$. The Nagelkerke R-square is 0.56.

TABLE 4, continued
Basis of Selection

[a]Participants were asked the following before producing puzzles for the paid portion of our experiment, "Relative to others, how would you gauge your potential to produce a large ***quantity*** of rebus puzzles?" They responded using a nine-point Likert-type scale with "1" being "much worse than others", 5 being "about the same as others", and 9 being "much better than others."

[b]Participants were asked the following before producing puzzles for the paid portion of our experiment, "Relative to others, how would you gauge your potential to produce a highly **creative** rebus puzzles?" They responded using a nine-point Likert-type scale with "1" being "much worse than others", 5 being "about the same as others", and 9 being "much better than others."

[c]Participants were asked the following before producing puzzles for the paid portion of our experiment, "How would you gauge your **quantity** potential relative to your **creativity** potential?" They responded using a nine-point Likert-type scale with "1" being "much better at quantity", 5 being "about the same at both", and 9 being "much better at creativity."

[d]Participants made fifteen choices between receiving a certain amount of $5.00 and participating in a lottery. Each lottery consisted of a chance of winning $10 with a probability of p and $0 with a probability 1-p. The probability of winning $10 in the first lottery was 85 percent. We decreased the chance of winning p in subsequent lotteries in 5 percent increments such that the chance of winning $10 in the final lottery was 15 percent. We recorded the number of the choice where each individual preferred the certain amount of $5.00 to the lottery such that the higher the score on this measure the higher the risk tolerance of participants.

TABLE 5
ANOVA for the Effect of Contract Type and Contract Determination on Highest Rated Puzzle

Panel A: Means (Standard Deviation) for Highest Rated Puzzle

	Contract Type	
	Quantity-Only	Creativity-Weighted
Contract Determination		
Random Assignment	7.33 (0.88)	7.22 (0.78)
Selection	6.91 (0.63)	7.26 (0.54)

Panel B: Analysis of Variance

Factor	*df*	*Sum of Squares*	*F*	*p-value (two-tailed)*
CONTRACT TYPE	1	0.35	0.68	.41
CONTRACT DETERMINATION	1	0.79	1.53	.22
CONTRACT TYPE × DETERMINATION	1	1.19	2.30	.13
Error	86			