Measuring and Motivating Quantity, Creativity, or Both

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

Bernhard E. Reichert
Doctoral Student
University of Texas at Austin
bernhard.reichert@phd.mccombs.utexas.edu

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

May 2007

We gratefully acknowledge funding from the Eugene and Dora Bonham Fund, the University of Texas University Research Institute, and a McCombs School of Business Research Excellence Grant. We received helpful comments on previous drafts from Robert Freeman, Michael Gibbins, Jeffrey Hales, Jessen Hobson, Karim Jamal, Bill Kinney, Lisa Koonce, Volker Laux, Sean Peffer, Shankar Venkataraman, Lynda Walker, Heather Wier, Jen Winchel, David Woods, and from accounting workshop participants at the University of Alberta and the University of Texas at Austin. We also appreciate the technical assistance provided by Morrie Schulman of the University of Texas at Austin Division of Instructional Innovation and by Todd Pinney of Turning Technologies, Inc. for providing and tailoring the application of radio-frequency response devices for use by our creativity raters.

Measuring and Motivating Quantity, Creativity, or Both

Abstract: We examine how worker productivity differs when performance-based compensation is based on measures of quantity, creativity, or the product of both measures. In an experimental task in which participants design "rebus puzzles," we find that quantity-based compensation increases the number of puzzles produced, and that creativity-based compensation improves average creativity ratings, as evaluated by an independent panel of raters. However, a weighted compensation scheme that rewards the product of quantity and average creativity ratings results in weighted productivity scores that are significantly lower than those generated by participants with quantity incentives alone. Follow-up analysis indicates that relative to participants compensated solely for quantity, participants in the weighted condition produce approximately the same number of high-creativity puzzles, but produce significantly fewer puzzles of mediocre creativity. This finding is consistent with the premise that participants rewarded for creativity-weighted output simplify their objective by restricting their production to high-creativity ideas, but are unable to translate this focus into a greater volume of high-creativity output. Implications address a possible explanation for why firms are reluctant to incorporate creativity measures within multi-dimensional performance measurement systems, notwithstanding published suggestions to do so.

Measuring and Motivating Quantity, Creativity, or Both

1. Introduction

A rich tradition of laboratory research in management accounting has examined the effects of performance-based compensation incentives on output measures (e.g., see reviews by Bonner et al. [2000]; Bonner and Sprinkle [2002]). This research stream has generally been restricted to objectively measurable outputs, such as the quantity produced. At the same time, "softer" dimensions of productivity, a prominent example of which is *creativity*, have been the focus of much attention in both the popular business press (e.g., Fallon and Senn [2006]) and in management and psychology (e.g., Amabile [1996], Eisenberger and Rhoades [2001]), but these efforts have generally not examined the same types of performance-contingent incentive structures as found in the accounting literature. Towards the goal of integrating these different perspectives, the current study examines the effects of performance-based compensation schemes that are contingent on explicit measures of quantity, creativity, or both.

A joint focus on quantity and creativity in performance-based compensation is useful because to succeed in an economy that is increasingly characterized by technological change and competitive threats to established product lines, firms must generate creative innovations while also maintaining high ongoing productivity (Chang and Birkett [2004]). To achieve these goals, Kaplan and Norton [1992, 1996] suggest incorporating measures of creativity and innovation in a balanced scorecard, along with more traditional performance measures. Yet, Ittner, Larcker, and Rajan [1997, p. 239] report from their survey of actual multi-dimensional bonus contracts that "despite calls for greater emphasis on innovation in performance evaluation (e.g., Kaplan and Norton [1992]), relatively few of the firms with non-zero weights on non-financial measures explicitly reported the use of new product development (6.1%) or innovation (2.6%) measures in their annual bonus formula." One reason why firms might be reluctant to reward employees for creativity is that creativity is difficult to measure. We control for this reason in an experimental setting that affords explicit creativity measures with reasonable interrater reliability, thereby allowing us to focus on more substantive issues involving the tradeoffs people make between striving to be creative and maximizing volume. To our knowledge, ours is the first study to examine how people respond to creativity incentive payments both in the presence and in the absence of simultaneous quantity incentives.

We examine these issues by asking experimental participants to design “rebus puzzles,” a task that affords meaningful measures of both quantity and creativity, with the latter dimension evaluated by independent raters who are blind to identities and treatment conditions. Both the raters and the experimental participants receive the same explicit definition of “creativity” as puzzles that are “original ideas, innovative, and clever.” In a 2 × 2 experimental design that calibrates pay formulas to hold *average* cash payments constant across cells, we manipulate whether participants’ compensation is fixed, based on the quantity produced, based on average creativity ratings achieved, or based on total creativity-weighted output (i.e., the product of quantity × average creativity).

We find that quantity-based compensation improves quantity and that creativity-based compensation improves average creativity ratings. However, defining “weighted productivity” as the sum of the creativity ratings for all output produced (which arithmetically equals the product of quantity and the average creativity rating of that quantity), we find that participants compensated to maximize weighted productivity fare significantly worse on this measure than participants compensated only for quantity. Follow-up analysis indicates that compensating participants for both creativity and quantity does not increase the number of high-creativity (i.e., overall top quartile) puzzles beyond those produced by participants with quantity incentives only, but does significantly reduce the production of “mediocre” puzzles (i.e., puzzles not in the overall top quartile). The reduction in mediocre output improves average creativity ratings as a “denominator effect,” while simultaneously lowering weighted productivity scores.

From a theoretical perspective, our findings are consistent with cognitive research indicating that people tend to simplify multi-dimensional objectives by prioritizing on one objective over others (Payne, Bettman, and Johnson [1993, Ch. 2]). For our task, participants compensated for creativity-weighted productivity likely adopt the simplified objective to produce as many high-creativity puzzles as possible. Their prioritization on creativity substantially impairs the production of puzzles that earn mediocre creativity ratings. Nevertheless, these participants do not generate an increased volume of high-creativity puzzles. The latter finding supports Amabile’s [1996] assertion that creativity does not emerge simply from trying harder, but with a more powerful test that evidences what people sacrifice when striving to be creative.

The practical implications of our findings for the design of multi-dimensional performance measures hinge on whether output of mediocre creativity is valued positively or negatively at the margin. If all output contributes positively, our results clearly indicate that participants with weighted quantity and creativity incentives sacrifice too much quantity for no appreciable gains in high-creativity output, suggesting that adding creativity to multi-dimensional performance measures could do more harm than good. However, for business situations that are better off without mediocre output than with it, our results suggest that rewarding both quantity and creativity serves as an effective screen to filter out mediocre efforts.

Section 2 develops our research questions for the quantity, creativity, and weighted productivity effects of individual or joint quantity and creativity incentives. Section 3 describes the experimental task and design, followed by results in Section 4 and conclusions and implications in Section 5.

2. Research Questions

In this section we develop two-tailed research questions for each of the three production dimensions we examine: quantity, creativity, and the weighted productivity of quantity × average creativity. Our preference for two-tailed research questions over directionally predicted hypotheses does not reflect lack of theory so much as it reflects alternative theories with potentially conflicting positions.[1] Put simply, the question of whether and when performance-based incentives are effective has yet to be resolved, especially in creative design tasks such as ours. Below we present both sides of the debate as applied to our dependent measures. We are ultimately interested in the joint effects of quantity and creativity incentives on creativity-weighted productivity (our third research question below), building incrementally to this question by first considering the effects of quantity and creativity incentives on separate measures of quantity and creativity.

2.1 RESEARCH QUESTION 1: WHAT ARE THE EFFECTS OF PERFORMANCE-BASED INCENTIVES ON QUANTITY?

2.1.1 RQ1a: What is the effect of quantity incentives on quantity? Before turning to creativity, we first consider the more basic effect of quantity incentives on the quantity produced. The traditional agency-theoretic presumption is that absent performance-contingent incentives, effort-averse agents will exert minimal effort and produce minimal output, such that the effect of explicit quantity incentives should be positive (Baiman [1990], Prendergast [1999]). Consistent with such arguments, experiments such as Chow [1983] and

---

[1] Equivalently, one can interpret our research questions as tests of null hypotheses against two-tailed alternative hypotheses, where one alternative stems from the agency-theoretic logic that performance-based incentive compensation should be effective for the dimension rewarded, while the other alternative presents a psychological premise for why performance-based incentives could actually do more harm than good.

Bailey, Brown, and Cocco [1998] have reported significant quantity gains resulting from piece-rate quantity incentives.

Even for simple quantity measures, however, the effectiveness of performance-based incentives is far from universally accepted, as reflected in the diverse findings reported in literature reviews by Camerer and Hogarth [1999], Bonner et al. [2000], and Bonner and Sprinkle [2002]. A common theme from these reviews is that incentive effects are task and setting specific. A potentially important task feature in the current study is the relatively open-ended and unstructured nature of designing “rebus puzzles,” raising the possibility that *task complexity* might negate the positive quantity incentive effects observed in other studies with more structured, algorithmic tasks (see Bonner and Sprinkle [2002], §3.2.1).

Beyond task complexity, it is also possible that participants might *enjoy* designing rebus puzzles, especially in an experimental setting in which participants have no practical alternative uses for the time scheduled, once they show up. Some studies claim that if extrinsic motives such as performance-based compensation undermine the intrinsic motivation from doing an enjoyable task, quantity incentives can actually prove harmful at the margin (e.g., Deci, Koestner, and Ryan [1999], Fessler [2003]).

The current study does not attempt to resolve or reconcile the divergent views and findings on the features that condition when quantity incentives are effective. We examine research question RQ1a simply to establish a baseline result for quantity that we can build upon in considering the incremental influence of incentives on creativity and the weighted product of quantity and creativity. That is, by examining quantity first, we control for various task-specific features that could influence *any* incentive effects in our experimental setting.

2.1.2 RQ1b: What is the effect of creativity incentives on quantity?

2.1.3 RQ1c: Do the effects of quantity and creativity incentives on quantity interact?

Any effect of creativity incentives on *quantity* likely depends on whether participants are motivated to increase quantity in the absence of creativity incentives. Accordingly, the answers to RQ1b (main effect of creativity incentives on quantity) and RQ1c (interaction of quantity and creativity incentives on quantity) are likely linked. If agency theoretic assumptions hold and quantity incentives exert a positive effect on quantity, then creativity incentives could reduce quantity as a competing demand when quantity incentives are present. Conversely, when quantity incentives are absent, participants would presumably have little if any motivation to produce, such that creativity incentives would do little additional harm to quantity at the margin.

While these arguments suggest the possibility of an interaction between the effects of quantity and creativity incentives on the quantity produced, any of the counterarguments raised earlier in the discussion of RQ1a (e.g., task complexity; intrinsic motivation) could also influence when and how creativity incentives influence quantity. As with RQ1a, we view the effects of incentives on *quantity* as an empirical question, establishing a baseline for our incremental consideration of creativity and creativity-weighted productivity.

2.2 RESEARCH QUESTION 2: WHAT ARE THE EFFECTS OF PERFORMANCE-BASED INCENTIVES ON AVERAGE CREATIVITY?

2.2.1 RQ2a: What is the effect of quantity incentives on average creativity? If quantity incentives motivate participants to produce as much volume as possible, then quantity incentives could result in lower average creativity ratings. This reasoning could hold both in the presence of creativity incentives, in which case quantity incentives serve as a competing demand, and in the absence of creativity incentives, in which case any inclination among participants to have fun with the exercise and design a few interesting, creative puzzles could be challenged by the incentive to crank out high volume. This reasoning hinges, of course, on the degree of creativity exhibited by participants both with and without creativity incentives, as we consider next.

2.2.2 RQ2b: What is the effect of creativity incentives on average creativity?

Research question RQ2b has been hotly debated in psychology, to the extent that Eisenberger and Rhoades [2001, p. 728] refer to it as the "reward-creativity controversy." One side of this controversy (to which Eisenberger and Rhoades [2001] themselves subscribe) holds that creativity incentives are generally effective in bringing about the desired result. One can think of this view as extending the traditional agency-theoretic assumption of effort-aversion to the domain of creativity, such that tangible rewards motivate more creative efforts. For example, Eisenberger and Rhoades [2001, pp. 733-734] asked college students to provide five possible titles for a short story that they were asked to read at the end of a scheduled class, with students in a randomly assigned treatment condition given the additional instruction that "if your titles are judged to be among the top half of the students in this class in terms of creativity, you will receive a financial reward next week for you to keep." Independent raters found the resulting titles to be more creative than those in a control condition without this incentive.

Other scholars in psychology and management assert that rewards for creative performance do more harm than good. Pivotal to this alternative view is that creativity differs substantively from other performance dimensions. Specifically, creativity requires imagination and insight, skills that are not necessarily responsive to external motives to work harder. A prototypical explanation of this view is well-captured by Amabile [1996, p. 153]:

> Unlike most desirable behaviors that psychologists study, creative behavior cannot be achieved simply by trying. Even individuals who have previously distinguished themselves for outstanding creativity often fail to produce creative work, despite their best efforts. Indeed, these individuals – for example, writers suffering "writer's block" – often complain that the harder they try, the more meager their success.

The argument continues that because creativity is unresponsive to trying harder, extrinsic rewards are doomed to fail. In an experiment predating Eisenberger and Rhoades [2001] but using a similar task, Kruglanski, Friedman, and Zeevi [1971] found that students promised a reward for generating possible titles for a paragraph produced titles that were judged to be significantly *less* creative than those in a control condition with no reward. In reconciling these apparently contradictory findings, Eisenberger and Rhoades [2001] observe that Kruglanski, Friedman, and Zeevi [1971] and more recent studies with similar results fail to explicitly tie rewards to creative output, such that participants might well misinterpret the rewards as an incentive to just "get the job done" rather than as an incentive to be creative. Thus, as Eisenberger and Rhoades [2001] demonstrate, incentives that are explicitly tied to creativity can indeed lead to more creative output. Yet, even Eisenberger and Rhoades [2001] do not fully test the implications of their premise, as their study manipulates creativity incentives but not quantity incentives. In contrast, we conduct the full 2 × 2 design, manipulating both performance-based quantity rewards and performance-based creativity rewards.

2.2.3 RQ2c: Do the effects of quantity and creativity incentives on average creativity interact? We have no *ex ante* basis for expecting an interaction between the effects of quantity and creativity incentives on average creativity, as we know little about the tradeoffs people make when faced with the possibility of quantity incentives, creativity incentives, or both. In the only previous effort of which we are aware that varied aspects of both quantity and creativity within the same experimental design, Shalley [1991] conducted a *goal-setting* study that manipulated the presence or absence of quantity and creativity goals such as a request to "do your best," but without implementing performance-based compensation for either dimension. Shalley [1991] found that the effects of quantity and creativity goals are additive. Specifically, quantity goals led to a greater number of proposed solutions to a series of business dilemmas, and creativity goals led to solutions that independent raters judged to be more creative, with no apparent sensitivity of either goal to the presence or absence of the other goal.[2] Bonner and Sprinkle [2002] are careful to distinguish goal-setting studies such as Shalley [1991] from incentive-compensation studies such as ours. Accordingly, we test for the possibility of interaction effects on average creativity ratings as an empirical question.

2.3 RESEARCH QUESTION 3: WHAT ARE THE EFFECTS OF PERFORMANCE-BASED INCENTIVES ON THE WEIGHTED PRODUCTIVITY OF QUANTITY × AVERAGE CREATIVITY?

Our primary contribution to the literature is that of explicitly measuring and manipulating incentives for weighted productivity, which we define as the sum of all creativity-weighted output, or equivalently, the product of quantity and the average creativity rating of that quantity. Although Shalley's [1991] goal-setting study manipulated both quantity and creativity goals, that study did not measure or manipulate incentives to maximize any joint function of both quantity and creativity. In our experiment, by contrast, participants in the experimental cell with both quantity and creativity incentives are compensated for maximizing the *multiplicative product of both measures*, such that optimal tradeoffs between quantity and creativity become critical. To the extent that businesses need both a prolific workforce and creative ideas from that workforce, our third research question is the most important question we consider.

2.3.1 RQ3a: What is the effect of quantity incentives on weighted productivity?

2.3.2 RQ3b: What is the effect of creativity incentives on weighted productivity?

Insofar as our measure of weighted productivity is strictly increasing in both quantity and creativity, one might expect that the main effects observed for quantity incentives on quantity (RQ1a) and for creativity incentives on average creativity (RQ2b) should also be manifest in the main effects of either incentive on weighted productivity. This reasoning seems most compelling when one incentive is present and the other is absent. When faced with both quantity and creativity incentives, however, workers must address the incremental challenge

---

[2] A follow-up study by Shalley [1995] found that creativity goals reduce quantity, but Shalley [1995] did not manipulate quantity goals and hence did not investigate tradeoffs between quantity and creativity.

of making tradeoffs between potentially conflicting motives to maximize a joint production measure. These tradeoffs suggest the possibility of an interaction effect, as we consider next.

2.3.3 RQ3c: Do the effects of quantity and creativity incentives on weighted productivity interact? If quantity and creativity incentives are individually effective in improving quantity and average creativity, respectively, it would be reasonable to expect some synergy from combining both dimensions, such that participants paid to maximize weighted productivity would be the most successful in doing so. This reasoning would support an interaction effect in which both quantity and creativity incentives are effective in increasing weighted productivity, but the combination of both incentives exceeds the sum of the individual main effects.

An important limitation to the above reasoning is the presumption that participants are able to make optimal tradeoffs between quantity and creativity when faced with both incentives. Quantity and creativity are characterized by fundamentally different features that could influence such tradeoffs. Quantity is objectively observable in real time, and is arguably more directly responsive to raw effort. Creativity, by contrast, is inherently subjective in evaluation and less clearly responsive to effort. The different nature of creativity relative to quantity could skew participants' attention to one dimension over the other. From an agency-theoretic perspective, risk-averse decision makers direct less effort towards noisier measures, *ceteris paribus* (Holmström and Milgrom [1991]), such that participants faced with both quantity and creativity incentives might focus primarily on quantity.

While agency theory suggests that participants compensated for weighted productivity could emphasize quantity over creativity, a more psychological perspective suggests a different possibility. Namely, several cognitive studies indicate that when presented with potentially conflicting motives, "people find making explicit tradeoffs emotionally uncomfortable, … not only because they are difficult to execute (cognitive effort) but also because they require the explicit resolution of difficult value tradeoffs" (Payne, Bettman, and Johnson [1993, p. 30]). A typical response to tradeoff aversion is a more noncompensatory model, in which one criterion becomes a qualifying condition that must be reached before another criterion is considered. A natural conditioning criterion in our weighted quantity × creativity condition is that puzzles must be of high creativity.[3] Viewing the task as "produce as many high-creativity puzzles as possible" is a heuristic representation of a more complex objective, consistent with Krishnan, Luft, and Shields' [2005] reasoning that people tend to simplify the task of weighting multiple performance measures by invoking incomplete mental models in lieu of formal analyses.

If participants in the weighted condition mentally simplify the weighted-compensation scheme by equating "maximize creativity-weighted output" with "produce as many high-creativity puzzles as possible," participants could overlook the productivity gains from more mediocre efforts when high-creativity ideas are unavailable. Moreover, if Amabile [1996] is correct that it is difficult to produce more creative efforts simply by trying harder, the strategy of maximizing high-creativity output might not be successful. Thus, creativity-weighted productivity incentives could lower the production of mediocre and low-creativity puzzles without increasing the production of high-creativity puzzles. This reasoning is a more subtle form of Amabile's [1996] assertion that creativity is not responsive to trying harder, but with a more powerful test that measures what people sacrifice when faced with conflicting motives to maximize both quantity and creativity.

---

[3] By way of analogy, Payne, Bettman, and Johnson [1993, p. 30] cite evidence from a study by Gregory et al. [1991] that people find it difficult to balance the benefits of fiscal responsibility and environmental protection, leading to hesitance to favor any initiatives that could harm the environment, irrespective of economic benefits.

3. Method and Design

3.1 PARTICIPANTS

We recruited 80 undergraduate business student volunteers to participate in one of eight compensated research sessions (10 per session), from which we obtained 78 usable participants.[4] We randomly assigned two sessions to each of four experimental conditions, as described shortly, yielding 18-20 participants per cell. Participants in each session worked independently at separate tables in a research laboratory,[5] such that we have no reason to expect session-specific effects, nor did we detect any such effects when comparing the two sessions within each experimental condition.

3.2 INSTRUCTIONS AND TASK

Experimental instructions informed participants that they would construct "rebus puzzles," defined as "a kind of riddle in which words and/or diagrams are used to represent a familiar term or phrase." While sometimes encountered in popular entertainment (e.g., Morris [1983]), we are unaware of any academic research applications of rebus puzzles other than an educational psychology application by Griggs [2000] to illustrate psychometric testing. Unlike Griggs [2000], our participants *design* rather than solve rebus puzzles. As shown in the Appendix, the experimental instructions common to treatment conditions gave participants eleven examples of rebus puzzles and corresponding solutions (mostly adapted from Griggs [2000]), emphasizing that rebus puzzles can be of a wide variety of types.

Also common to all experimental conditions was the following wording: "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

---

[4] One volunteer did not show up, and one volunteer did not follow the instructions to indicate the solutions to his/her rebus puzzles on the backs of the cards, thereby precluding creativity evaluations. Because the experiment involves individual actions, the slightly unbalanced cell sizes should not pose any problem.

[5] We did not even allow oral questions, guarding against any session-wide influences beyond the experimental instructions. Participants with a question were instructed to write the question down on paper, upon which one of the experimenters wrote an answer to the question on the same paper at the participant's individual table.

puzzles (i.e., puzzles that are original ideas, innovative, and clever)." Thus, all participants were informed of the experimenters' desire for both quantity and creativity. In this manner, all of our experimental conditions capture explicit goals similar to the combination of "do your best" quantity and creativity goals in Shalley's [1991] (uncompensated) goal-setting study, such that we can focus the current research on the incremental incentive effects of performance-contingent compensation measures (Bonner and Sprinkle [2002]).

Participants designed their puzzles on 3 × 5 inch index cards, placing the puzzle on one side and its solution on the other side. Participants had 20 minutes to design puzzles, putting each completed puzzle design in an "output box," subject to the understanding that once in the box, a puzzle could not be removed. The last part of the instructions before beginning this task was the explanation of the compensation participants would receive from the experiment.[6] This paragraph was the only part of the instructions that varied across conditions, as explained next.

3.3 EXPERIMENTAL DESIGN AND TREATMENT CONDITIONS

Our study implements a 2 × 2 between-subjects design, manipulating quantity incentives and creativity incentives. In the control condition with neither quantity nor creativity incentives, the instructions informed participants that each would receive a $25 fixed payment in approximately two weeks, "no matter what you do today." The instructions continued to explain (truthfully) that the only reason for waiting two weeks before distributing payments is that "different versions of the research require waiting, and we want to pay all participants at the same time." Thus, the control condition holds constant the *presence* of cash payments and the delay in those payments necessitated in other conditions (see below), but removes any performance-contingent element of participants' compensation.

---

[6] As a pre-task comprehension check, participants were asked to provide a brief written description of how they would be compensated. One of the experimenters checked these answers before continuing.

In the condition with quantity incentives only, the instructions informed participants that their compensation to be paid in approximately two weeks would "be based on *how many* puzzles you can construct," using a payment rate per puzzle to result in $5.00 for the participant (among all participants with that version of the instructions) submitting the fewest puzzles and $45.00 for the participant submitting the most puzzles, thereby yielding an average compensation of approximately $25.00. Importantly, while the minimum and maximum anchors are based on actual performance, this payment scheme and the similar creativity and weighted productivity schemes described below are not "tournament" schemes in the sense described by Bonner et al. [2000]. That is, in the current scheme, producing more puzzles results in greater compensation, as the instructions emphasized, irrespective of ordinal ranking. The advantage of using actual performance to determine the minimum and maximum anchors for the linear payment rate is that we could pre-commit to an average payment around $25.00, thereby holding the *average magnitude* of compensation constant while manipulating the *nature* of that compensation.

In the condition with creativity incentives only, the instructions informed participants that their compensation would be based on the average creativity ratings on a 1-to-10 scale awarded to their puzzles by doctoral-student raters. The instructions noted that average creativity is simply the sum of the individual creativity ratings divided by the number of puzzles submitted. As with the quantity scheme explained above, the instructions explained that the payment rate would be anchored to result in $5.00 for the participant (within that version) with the lowest average creativity rating and $45.00 for the participant with the highest average rating, thereby resulting in average compensation around $25.00.

In the condition with both quantity and creativity incentives, the instructions informed participants that their compensation would be based on "the creativity-weighted total score of all rebus puzzles you can construct in 20 minutes," determined by adding the 1-to-10 creativity ratings as evaluated by doctoral-student raters for all puzzles submitted. The instructions emphasized the summation of creativity ratings to clarify that “each puzzle you submit helps your total score,” while also pointing out that “higher rated puzzles count more.” That said, weighted productivity can also be obtained by multiplying quantity by the average creativity rating of that quantity, such that the measure reduces to a multiplicative combination of the quantity and average-creativity conditions. Similar to the other performance-contingent compensation conditions, the instructions explained that the payment rates would result in $5.00 and $45.00 for the lowest and highest creativity-weighted total scores, respectively, yielding an average payment around $25.00.

After finishing the task, participants completed a post-experimental questionnaire and were dismissed with a request not to discuss the experiment with others. As promised, about two weeks later, we determined the payment rates specified in the instructions and distributed cash payments privately at a location communicated by email to all participants. Payments ranged from $5.00 to $45.00, as indicated in the instructions, without deception of any form.

3.4 DETERMINING CREATIVITY RATINGS

To obtain the dependent variable for creativity and to implement our creativity-based payment schemes, we needed unbiased ratings of creativity. As indicated in the instructions, we used business doctoral students for this purpose. Initially, after assigning unobtrusive identification codes to all 1,360 rebus puzzles submitted by the 78 participants, we shuffled these cards to expedite individual ratings by two doctoral students who were blind to identities and treatment conditions, thereby enabling us to calculate compensation formulas and pay participants on a timely basis. Each of the two raters simply placed the cards into ten stacks, ranging from 1 (lowest creativity) to 10 (highest), after which we tabulated the ratings and reshuffled the cards for the next rater.

Amabile [1996, Ch. 3] advises using several raters in creativity research, both to reduce noise and to ensure reliability. Accordingly, after the initial evaluations described above, we constructed a database of scanned images and solutions for the 1,360 puzzles, allowing us to project each puzzle on computer screens for evaluation by nine additional doctoral-student raters. These additional raters used radio-frequency response keypads to evaluate the puzzles concurrently (but independently) in a rating session that took about six hours. As with the initial two raters, the nine subsequent raters viewed puzzles in random order, and were blind to treatment conditions. We obtain the same statistical conclusions for testing our research questions whether we include or omit the initial two raters who followed a somewhat different protocol, so we report averages for all eleven raters as our primary measure of creativity. All raters read the same common instructions the participants had, without the final paragraph about compensation that varied by treatment condition. All eleven doctoral-student raters served as compensated research assistants for this exercise, and were otherwise independent of the study.

Correlations of average creativity ratings between each doctoral student rater and the mean rating of the other ten raters are positive for all raters, ranging from 0.05 to 0.84. Such correlations are statistically significant for ten of the eleven raters, and are above 0.60 for eight of the eleven. As a measure of interrater reliability, Cronbach's Alpha is 0.86, which exceeds typical reliability thresholds (Peterson [1994]). We conclude that our raters are reasonably consistent, such that their average is a reliable measure of creativity. To verify the robustness of our findings to an even more reliable creativity measure, we later report a supplemental analysis using averages from only the eight raters with the highest interrater correlations.

4. Results

We use analysis of variance (ANOVA) to investigate the main and interactive effects of quantity and creativity compensation incentives on production quantity, average creativity ratings, and the weighted productivity of quantity × average creativity. We also conducted various analyses of covariance (ANCOVA), using responses from a post-experimental questionnaire as covariates to control for self-perceptions of familiarity with rebus puzzles, creative ability, and the like. These covariates do not change our statistical conclusions, so for simplicity we report the ANOVA results as our primary findings.[7]

4.1 EFFECTS OF PERFORMANCE-BASED INCENTIVES ON QUANTITY

Our first research question addresses the effect of incentives on quantity. Descriptive statistics reported in Table 1, Panel A indicate that quantity is highest among participants with quantity incentives only (average of 28.6 puzzles produced), followed by participants with both quantity and creativity incentives (17.5). Participants in the fixed and creativity-only conditions produce the fewest puzzles (11.8 and 11.3, respectively).

Table 1, Panel B reports an ANOVA with quantity as the dependent measure, showing a significant positive main effect of quantity incentives on quantity (RQ1a), a significant negative main effect of creativity incentives on quantity (RQ1b), and a significant interaction effect (RQ1c). As is visually apparent in Figure 1, follow-up analysis of the interaction effect in Table 1, Panel C reveals that the negative effect of creativity incentives on quantity holds only when quantity incentives are present. This interaction qualifies the main effect of creativity incentives detected in RQ1b.

The results for the effects of incentives on quantity largely support agency-theoretic assumptions. In the absence of quantity incentives, participants produce relatively few

---

[7] We also tested for gender differences, finding no effect of participant gender on creativity ratings.

puzzles, whether or not they have creativity incentives. Quantity incentives prompt a significant jump in output, but that jump is muted when creativity incentives are also present as a competing demand. Given this baseline, we now turn to the effect of incentives on creativity.

4.2 EFFECTS OF PERFORMANCE-BASED INCENTIVES ON AVERAGE CREATIVITY RATINGS

Consistent with the descriptive statistics shown in Panel A of Table 2 and depicted in Figure 2, the ANOVA in Panel B of Table 2 indicates a significant negative main effect of quantity incentives on average creativity ratings (RQ2a), a significant positive main effect of creativity incentives on average creativity (RQ2b), and no discernable interaction (RQ2c).

The statistically significant positive effect of creativity incentives on average creativity is consistent with the agency-theoretic premise that incentives lead to better performance for the dimension that is measured and rewarded. Apparently our participants can recognize efforts of higher creativity, generating a higher proportion of creative puzzles when creativity is rewarded explicitly. However, we emphasize that average creativity ratings reflect both the numerator of total creativity scores and the denominator of total quantity, such that one way to improve an average is simply to lower the denominator. As we report shortly in the consideration of weighted productivity, this “denominator effect” accounts for much of the improvement in average creativity ratings attributable to creativity incentives.

Quantity incentives reduce average creativity ratings whether creativity incentives are present or absent, such that we do not detect an interaction between quantity and creativity incentives on average creativity ratings. One reason why quantity incentives might lower creativity ratings even in the absence of creativity incentives is that participants in the fixed-pay control condition produce a small number of puzzles, but have some fun with the exercise by making those puzzles creative enough to alleviate boredom (essentially an intrinsic incentive). Quantity incentives prompt a larger volume of cards with a wider distribution of creativity ratings, as we explore in more detail later.

4.2.1 What is creativity? A fundamental limitation of creativity research is that the construct of interest is a subjective matter of taste, calling into question what exactly our raters are capturing when they rate one puzzle as being more creative than another. Fortunately, an advantage of our task is that rebus puzzles can be classified into a variety of patterns, allowing us to gain some insight into what participants and raters thought was "creative." Specifically, the instructions (see Appendix) provided participants with eleven examples of rebus puzzles. These examples could serve as prompts to generate other ideas along the same themes. For example, the illustration in the instructions of "man overboard" (the word "man" written above the word "board") could prompt several other "something over something" puzzles. Creativity incentives, by contrast, could motivate the desire to do something different, consistent with the instructional definition of creativity as puzzles that are "original ideas, innovative, and clever."

To test this reasoning, we coded all puzzles as ideas that did or did not follow the same pattern as one of the instructional examples.[8] We then determined each participant's percentage of puzzles that followed such patterns. Table 3 reports descriptive statistics for these percentages in Panel A and a two-way ANOVA in Panel B. This analysis reveals a statistically significant negative main effect of creativity incentives on the percentage of puzzles patterned after an instructional example, a marginally significant *positive* main effect of quantity incentives (two-tailed p = .07), and no interaction. As depicted in Figure 3, these

---

[8] Our exact algorithm codes a puzzle as following a instructional pattern if it meets any of the following eleven criteria (words in quotation marks had to appear in the solution; words in parentheses are the solutions to the corresponding instructional examples): (1) something "over" something (man overboard); (2) something "under" something (I understand); (3) something "between" something (just between you and me); (4) something "in" something (gross injustice); (5) something "below" something (three degrees below zero); (6) puns based on the numbers 2 or 4 to represent "to," "too," or "for" (too funny for words); (7) words "growing" in size (growing pains); (8) crossed words (cross roads); (9) "high" words or pictures at the top of the card (high chair); (10) something with a hole or holes in it (a hole in one); or (11) something crossed out using the "not" symbol (to be or not to be).

findings suggest that our primary results for creativity ratings are inversely correlated with participants' propensity to submit puzzles similar to those provided as examples in the instructions.[9]

As illustrations, Panels A, B, and C of Figure 4 are three examples of highly rated puzzles from our participants, each uniquely different from any theme in the instructional examples. Conversely, Panel D of Figure 4 is a relatively low-rated puzzle (ranking 1,000 in creativity out of 1,360) that is a straightforward extension of the "something over something" theme illustrated in one of the instructional examples. Puzzles like the Panel D example likely facilitate a quantity-focused strategy, presuming that participants gain production efficiencies by extending the instructional examples to similar ideas (like extending an existing product line). We next integrate the creativity and quantity results by considering the multiplicative weighted productivity of both measures.

4.3 EFFECTS OF PERFORMANCE-BASED INCENTIVES ON THE WEIGHTED PRODUCTIVITY OF QUANTITY × AVERAGE CREATIVITY

Our third research question addresses the effect of incentives on weighted productivity, defined as the sum of the creativity ratings for all output produced, or equivalently, the product of quantity × average creativity. Table 4, Panel A reports and Figure 5 depicts mean weighted productivity scores by treatment condition. By far the highest weighted productivity scores occur in the cell with quantity incentives only (mean of 123). Participants in the cell with both quantity and creativity incentives score considerably lower (mean of 83), even though these were the only participants compensated for weighted productivity. The lowest weighted productivity scores occur in the cells with no incentives or creativity incentives only (means of 58 and 59, respectively). The ANOVA in Table 4, Panel B corroborates the pattern evident in Figure 5. We find a statistically significant

---

[9] Corroborating this assertion, the average creativity rating of 4.30 for puzzles patterned after an instructional example is significantly lower than the average rating of 4.86 for puzzles *not* patterned after an instructional example ($t = 8.48$; two-tailed $p < .01$).

positive main effect of quantity incentives on weighted productivity (RQ3a), a statistically significant *negative* effect of creativity incentives (RQ3b), and a statistically significant interaction (RQ3c). To follow up on the interaction, simple-effect tests in Table 4, Panel C reveal a statistically significant negative effect of creativity incentives on weighted productivity in the presence of quantity incentives, and no effect of creativity incentives in the absence of quantity incentives.

How is it possible that quantity incentives significantly improve quantity (RQ1), creativity incentives significantly improve average creativity ratings (RQ2), and yet the combination of quantity and creativity incentives fares significantly worse than quantity incentives only when computing the weighted product of both measures (RQ3)? Figure 6 provides much of the answer. Figure 6 divides the total output in each experimental cell into "high-creativity" puzzles, defined as puzzles with a consensus rating above 5.5 (approximately the top quartile of all puzzles), and all other puzzles. It reveals that all four experimental conditions are remarkably similar in the *absolute number* of high-creativity puzzles produced. However, quantity incentives lead participants to *also* produce a large number of mediocre puzzles rated below 5.5, while creativity incentives suppress such efforts. Thus, quantity incentives significantly increase total quantity, while creativity incentives significantly increase *average* creativity ratings by lowering the production of mediocre puzzles. Because output of any creativity strictly increases weighted productivity, participants with quantity incentives only dominate, insofar as they produce approximately the same volume of high-creativity output while also producing a substantial volume of less creative output.

To corroborate these inferences, we conduct three supplemental analyses. First, an ANOVA on the number of puzzles rated above 5.5 finds no significant main effects or interaction (Table 5), supporting our assertion that participants across cells achieved approximately the same volume of high-creativity output. Of course, one reason for the lack of a statistically significant difference could be that our measure of creativity is imprecise. Hence, to give creativity-weighted pay its “best shot,” we conduct a second supplemental analysis that constructs creativity rating averages from the eight raters whose correlations with the rest of the panel are 0.60 or higher. This construction increases interrater reliability (as measured by Cronbach’s alpha) from 0.86 to 0.93. Yet, we continue to find similar results for the number of puzzles with an average rating above 5.5, detecting no main effects or interaction at conventional significance levels.[10] Using the eight highest correlated raters in lieu of all eleven raters also generates the same statistical conclusions for all other analyses reported previously.

As a third supplemental analysis, we focus attention on the *highest* rated puzzle each participant produced. Whether using all eleven raters or the subset of eight with the highest interrater correlations, ANOVAs (not tabulated) detect no main or interaction effects of performance-based compensation.[11]

4.3.1 Discussion of weighted productivity results. The similar production of high-creativity output across cells implies that differences in weighted productivity scores are driven primarily by the production of puzzles that earn *mediocre* creativity ratings, as is apparent in Figure 6. The irony in this finding is that the participants in our weighted productivity condition could have improved their weighted productivity scores by ignoring

---

[10] While it is of insufficient magnitude to generate a statistically significant main effect or interaction, one possible exception to approximately equal high-creativity output across conditions occurs in the fixed-payment control condition. For analyses using eleven (eight) raters, participants in the fixed-payment condition produce an average of 3.39 (2.89) high-creativity puzzles, in comparison to an average of 4.48 (4.32) across the other three cells. Statistically, an *ad hoc* pairwise comparison between the control condition and an equally weighted composite of the other three cells is at least marginally significant (t = 1.65; two-tailed p = .10 for eleven raters, or t = 2.11; p = .04 for eight raters). While this result suggests that *any* incentive facilitates a reasonable production of high-creativity output, the more important point for our research conclusions is that the three incentive conditions do not differ from each other (p > .50 with eleven or eight raters).

[11] The average highest rating by participant is 6.80 (6.95) in the analysis with eleven (eight) raters. Neither main effect is statistically significant (two-tailed p = .14 for quantity incentives in both analyses and p > .50 for creativity incentives), nor is there any discernable interaction (p = .45 or p > .50 with eleven and eight raters, respectively).

their creativity incentives and emulating the production strategies of their counterparts with quantity incentives only. Hence, at the margin, creativity incentives impaired the ability of these participants to maximize the measure for which they were compensated.

As developed earlier in the discussion of RQ3c, a likely explanation for this finding is that participants compensated for weighted productivity simplified their objective by prioritizing on creativity rather than maximizing weighted-productivity scores *per se*. To corroborate this explanation, we analyze strength of agreement with two post-experimental statements, each elicited on a 1-to-7 Likert scale: (1) "I worked hard to construct creative rebus puzzles," and (2) "I worked hard to increase the number of rebus puzzles I constructed." The only statistically significant effect from an ANOVA on the *difference* between the creativity and quantity responses is the main effect of creativity incentives (two-tailed $p = .01$). Specifically, the mean difference score is positive (favoring creativity) only in the two cells with creativity incentives (mean difference of 0.78) as compared to the two cells without creativity incentives (mean difference of -0.55). Thus, participants perceived that they worked harder on creativity when they were paid for creativity. However, Figure 6 and Table 5 indicate that creativity incentives did not enable participants to produce a greater volume of high-creativity puzzles.

The fact that participants in the weighted productivity condition did not produce more high-creativity puzzles than participants in the quantity-only condition suggests that our weighted productivity findings reflect more than just a suboptimal tradeoff between quantity and creativity. A suboptimal tradeoff would imply that participants sacrifice too much in one dimension to gain too little in another dimension, but our weighted productivity participants appear to be sacrificing quantity without generating *any* compensating gains in high-creativity output. As such, while weighted productivity participants likely focused too much on creativity, the more fundamental problem is that this focus did not help these participants to be more creative (in an absolute sense).[12] While Amabile [1996] does not address this result directly, we believe that our findings are consistent with a more subtle form of Amabile's [1996] premise that working harder to be creative does not necessarily lead to greater creativity. In this manner, "creativity" likely differs from other aspects of production quality that are more amenable to improvement under multi-dimensional incentives (e.g., Farrell, Kadous, and Towry [2006]).[13] Our findings suggest that weighted-productivity participants became bogged down trying to generate high-creativity ideas. Conversely, quantity-only participants were producing puzzles of more mediocre creativity when high-creativity ideas were unavailable, but were also producing high-creativity puzzles when those puzzles emerged naturally from the production process.

The additional "mediocre" puzzles produced by participants in the quantity-only condition increased their weighted-productivity scores for testing RQ3, but were these puzzles of sufficient merit to be interpretable as output that truly improved "productivity"? We believe so, as our doctoral-student raters awarded an average creativity rating of 4.22 to the puzzles that we classify as mediocre (i.e., below 5.5), and this average does not differ significantly due to quantity incentives ($p > .50$), creativity incentives ($p = .46$), or the interaction between the two ($p > .50$). Thus, most "mediocre" puzzles were of modest, but not unreasonably low creativity, often based on instructional patterns as typified by the example in Panel D of Figure 4. To be sure, some puzzles made little sense, such as the example in Figure 4, Panel E of "Chewing gum," which received the lowest rating of all 1,360 puzzles. Puzzles of this nature were more the exception than the rule, and even the

---

[12] This finding mitigates any concern about the emphasis on creativity in the instructional wording, such as a clause in the weighted productivity condition observing that "at the extreme, a puzzle rated 10 counts ten times as much as a puzzle rated 1." Even if wording of this nature encouraged a noncompensatory, "creativity-first" strategy, it did not help participants to produce a greater volume of high-creativity puzzles. Certainly it is possible that a weaker operationalization of the weighted productivity condition to dilute the focus on creativity could have improved quantity, but that objective would seem better served by simply reverting to the quantity-only condition, in which participants maximize quantity with no apparent loss of high-creativity output.

[13] Participants in Farrell, Kadous, and Towry [2006] constructed virtual (computerized) sandwiches, in which "quality" was defined as a sandwich that matched the customer's order. Their task did not involve creativity.

participant who submitted “Chewing gum” also submitted the relatively highly rated and clever “Popeye” in Panel F of Figure 4. Hence, even if some mediocre output is without value, our results suggest that a “rebus firm” would be no worse off, and could potentially be better off, with quantity-only incentives, so long as mediocre output is not actually *harmful* at the margin.

5. Implications, Limitations, and Future Directions

In what, to our knowledge, is the first study to manipulate performance-contingent compensation for both quantity and creativity within the same experiment, we find that quantity incentives lead participants to produce significantly more rebus puzzles, and creativity incentives lead to puzzles that generate significantly higher average creativity ratings by independent doctoral-student raters. However, participants paid to maximize the weighted productivity of quantity × average creativity fare significantly worse in doing so than participants paid only to maximize quantity. A likely reason for this result is that participants in the weighted productivity condition prioritize on maximizing high-creativity output. This prioritization suppresses productivity gains from more mediocre puzzles when high-creativity ideas are unavailable, but does not appear to facilitate any compensating gains from an increased volume of high-creativity puzzles, consistent with Amabile’s [1996] premise that creativity does not emerge simply from trying harder. Thus, we find that creativity incentives improve average creativity ratings primarily as a “denominator effect,” reducing the number of mediocre puzzles without significantly increasing the number of high-creativity puzzles.

For practice, the implications of these findings hinge on whether mediocre output is valued positively or negatively at the margin. An analogy to academic research incentives serves to illustrate the point. Assume that university research reputations stem primarily from high-creativity contributions (Dewett and Denisi [2004]). Assume further that hypothetical

University A aligns faculty incentives with the goal of high-creativity output by placing much greater weight in merit review evaluations on high-creativity articles (perhaps as proxied by top-tier journals), while University B simply rewards faculty for publishing research articles (quantity). *Ceteris paribus,* our results suggest that faculty members at both universities are likely to produce a few top-tier articles, on average, whereas faculty members at University B are also likely to produce several other articles of lesser creativity. If University B's reputation is actually *harmed* at the margin by these less-creative contributions, B would be better off with a weighted incentive scheme. But if all research makes a positive contribution (albeit of varying degrees), our results suggest that B's simpler incentive structure could generate the benefits of much higher *overall* research volume without sacrificing the benefits of high-creativity volume. Of course, it would be difficult to test this premise in a world in which faculty members are not assigned randomly to universities. We return to the issue of selection bias shortly.

For management accounting, our results bear upon the role of creativity in multi-dimensional performance measures such as the balanced scorecard (Kaplan and Norton [1996]). Ittner, Larcker, and Rajan [1997] observe that despite calls for incorporating creativity-related measures in multi-dimensional performance compensation schemes, firms are reluctant to do so. Our findings suggest that this hesitance may reflect more than just difficulty in measuring creativity. Specifically, even if creativity is measurable and important to firm success, imbedding creativity measures in multi-dimensional evaluation and compensation schemes could have the unintended consequence of suppressing less-creative productivity without necessarily generating gains in high-creativity output. Again, whether that consequence is good or bad depends on whether output of lesser creativity is valued positively or negatively at the margin.

It is tenuous, of course, to base any of these conclusions on a single laboratory experiment in which undergraduate business students design rebus puzzles. Future studies could manipulate various environmental factors that could influence our results. For example, one of the primary decision-influencing roles of incentive compensation schemes is in mitigating the adverse-selection problem of hidden information, using selection into different compensation schemes to attract employees of different abilities (Demski and Feltham [1978], Sprinkle [2003, §2.1.1]). In our experiment, random assignment to treatment conditions allows us to separate the motivational effects of the incentive schemes from the effects of self-selection. However, this benefit also limits our ability to generalize results. It is plausible that if we had allowed participants to *choose* their incentive schemes, and if participants had sufficient self-insight, the weighted productivity condition might have attracted a greater proportion of relatively creative participants, mitigating the relative disadvantage of that condition among randomly assigned participants. We view the selection issues involving multi-dimensional performance measures as a prime avenue for further study, extending the single-dimension context of prior research in this area (as reviewed by Sprinkle [2003, §2.1.1]).

In addition to self-selection, two other important factors limit our results and could be explored in future research. First, we did not provide participants with feedback, and therefore do not address the role of feedback in moderating incentives (e.g., Sprinkle [2000]). Part of this limitation is realistic, insofar as real-time creativity feedback is inherently difficult to provide both in the laboratory and in practice, given the delayed evaluation of creativity. Moreover, it is unclear whether feedback would mitigate or exacerbate the effects we detect, as creativity feedback might lead weighted-productivity participants to fixate even more on creativity, further undermining their willingness to produce mediocre puzzles. Still, we acknowledge the lack of real-time feedback as a potential limitation. A final important limitation is that the incentives we implement in this research do not tap the potential for *budgets* to provide incremental motivation (Covaleski et al. [2003]). We favored linear incentive contracts in the spirit of taking one step at a time, given that we are unaware of any prior experimental study of performance-based compensation for both quantity and creativity. Any number of budgeting schemes and issues provide further opportunities for research on creativity incentives.

While psychology and organizational behavior scholars have written a great deal about “soft” performance attributes such as creativity, we believe that an accountant’s perspective can contribute significantly at the margin. As reviewed by Bonner et al. [2000], Bonner and Sprinkle [2002], Covaleski et al. [2003], and Luft and Shields [2003], among others, management accounting researchers have contributed significant insights and structure to several incentive issues, but most of these issues have involved relatively objective, verifiable performance measures. Although softer measures of performance such as creativity are inherently subjective, there is no conceptual reason why they cannot also be investigated within a management accounting framework, as we have attempted to illustrate in the current study.

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:

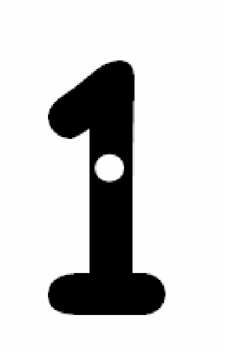
$\frac{\text{Man}}{\text{Board}}$ $\frac{\text{Stand}}{\text{I}}$ You Just Me   Just 144 ice

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. Use these cards to construct your own rebus puzzles, putting the ***puzzle on the front side and the solution on the back side***. 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).

You will have 20 minutes to construct your rebus puzzles.

Compensation

[Control condition with neither quantity nor creativity incentives:]

You will receive a ***fixed payment of $25.00*** for constructing rebus puzzles for 20 minutes. In about two weeks, all participants with this version of the research will receive $25.00 in cash. You will not need to do anything else, and you will get $25.00 no matter what you do today. The only reason for waiting two weeks is that different versions of the research require waiting, and we want to pay all participants at the same time. We promise that you and all others with this version of the research will receive a fixed payment of $25.00 cash for participating today.

[Condition with quantity incentives only:]

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 participating in this version of the research. We will then determine a cash payment per puzzle, where the payment rate per puzzle results in **$45.00** total compensation for the participant (or participants, if tied) who construct the ***most*** rebus puzzles, and **$5.00** for the participant (or participants, if tied) who construct the ***fewest*** rebus puzzles. Everyone else 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, 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 with this version of the research will receive cash compensation as described above for participating today.

[Condition with creativity incentives only:]

Your compensation will be based on the ***creativity*** of the rebus puzzles you can construct in 20 minutes. To determine this, we will ask a group of doctoral students 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 then calculate the simple ***average*** (i.e., mean) of your creativity ratings by dividing the sum of your creativity ratings by the number of cards you submit. That average will be the basis of your compensation. Specifically, we will determine a cash payment rate to result in **$45.00** total compensation for the participant (or participants, if tied) with the ***highest*** average creativity rating, and **$5.00** for the participant (or participants, if tied) with the ***lowest*** average creativity rating. Everyone else will receive something in between $5.00 and $45.00, depending on average creativity ratings, to result in an expected average compensation around **$25.00.** The higher your average creativity rating, 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 with this version of the research will receive cash compensation as described above for participating today.

[Condition with quantity and creativity incentives:]

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 doctoral students 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 then ***add*** the doctoral students’ creativity ratings on a 1 to 10 scale of ***all*** puzzles you submit. Thus, ***each puzzle you submit helps your total score, 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 determine a cash payment rate to result in **$45.00** total compensation for the participant (or participants, if tied) with the ***highest*** total score, and **$5.00** for the participant (or participants, if tied) with the ***lowest*** total score. Everyone else 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, 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 with this version of the research will receive cash compensation as described above for participating today.

REFERENCES

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

BAILEY, C. D.; L. D. BROWN; AND A. F. COCCO. "The Effects of Monetary Incentives on Worker Learning and Performance in an Assembly Task." *Journal of Management Accounting Research* 10 (1998): 119-131.

BAIMAN, S. "Agency Research in Managerial Accounting: A Second Look." *Accounting, Organizations and Society* 15 (1990): 341-371.

BONNER, S. E.; R. HASTIE; G. B. SPRINKLE; AND S. M. YOUNG. "A Review of the Effects of Financial Incentives on Performance in Laboratory Tasks: Implications for Management Accounting." *Journal of Management Accounting Research* 12 (2000): 19-64.

BONNER, S. E., AND G. B. SPRINKLE. "The Effects of Monetary Incentives on Effort and Task Performance: Theories, Evidence, and a Framework for Research." *Accounting, Organizations and Society* 27 (2002): 303-345.

CAMERER, C. F., AND R. M. HOGARTH. "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework." *Journal of Risk and Uncertainty* 19 (1999): 7-42.

CHANG, L., AND B. BIRKETT. "Managing Intellectual Capital in a Professional Service Firm: Exploring the Creativity-Productivity Paradox." *Management Accounting Research* 15 (2004): 7-31.

CHOW, C. W. "The Effects of Job Standard Tightness and Compensation Scheme on Performance: An Exploration of Linkages." *The Accounting Review* 58 (1983): 667-685.

COVALESKI, M. A.; J. H. EVANS III; J. L. LUFT; AND M. D. SHIELDS. "Budgeting Research: Three Theoretical Perspectives and Criteria for Selective Integration." *Journal of Management Accounting Research* 15 (2003): 3-49.

DECI, E. L.; R. KOESTNER; AND R. M. RYAN. "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation." *Psychological Bulletin* 125 (1999): 627-668.

DEMSKI, J. S., AND G. A. FELTHAM. "Economic Incentives in Budgetary Control Systems." *The Accounting Review* 53 (1978): 336-359.

DEWETT, T., AND A. S. DENISI. "Exploring Scholarly Reputation: It's More Than Just Productivity." *Scientometrics* 60 (2004): 249-272.

EISENBERGER, R., AND L. RHOADES. "Incremental Effects of Reward on Creativity." *Journal of Personality and Social Psychology* 81 (2001): 728-741.

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

FARRELL, A. M.; K. KADOUS; AND K. L. TOWRY. "Contracting on Contemporaneous vs. Forward-Looking Measures: An Experimental Investigation." Working paper, University of Illinois at Urbana-Champaign and Emory University, 2006.

FESSLER, N. J. "Experimental Evidence on the Links among Monetary Incentives, Task Attractiveness, and Task Performance." *Journal of Management Accounting Research* 15 (2003): 161-176.

GREGORY, R., H.; KUNREUTHER, D.; EASTERLING; AND K. RICHARDS. "Incentive Policies to Site Hazardous Waste Facilities." *Risk Analysis* 11 (1991): 667-675.

GRIGGS, R. A. "A One-Minute 'Intelligence' Test." *Teaching of Psychology* 27 (2000): 132-134.

HOLMSTRÖM, B., AND P. MILGROM. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics & Organization* 7 (1991): 24-52.

ITTNER, C. D.; D. F. LARCKER; AND M. V. RAJAN. "The Choice of Performance Measures in Annual Bonus Contracts." *The Accounting Review* 72 (1997): 231-255.

KAPLAN, R., AND D. NORTON. "The Balanced Scorecard—Measures that Drive Performance." *Harvard Business Review* 70 (1992): 71-79.

KAPLAN, R., AND D. NORTON. *The Balanced Scorecard: Translating Strategy Into Action*. Boston, MA: Harvard Business School Press, 1996.

KRISHNAN, R.; J. L. LUFT; AND M. D. SHIELDS. "Effects of Accounting-Method Choices on Subjective Performance-Measure Weighting Decisions: Experimental Evidence on Precision and Error Covariance." *The Accounting Review* 80 (2005): 1163-1192.

KRUGLANSKI, A. W.; I. FRIEDMAN; AND G. ZEEVI. "The Effects of Extrinsic Incentive on Some Qualitative Aspects of Task Performance." *Journal of Personality* 39 (1971): 606-617.

LUFT, J., AND M. D. SHIELDS. "Mapping Management Accounting Graphics and Guidelines for Theory-Consistent Empirical Research." *Accounting, Organizations and Society* 28 (2003): 169-249.

MORRIS, S. *Omni Games: The Best Brainteasers from Omni Magazine*. New York: Holt, Rinehart & Winston, 1983.

PAYNE, J. W.; J. R. BETTMAN; AND E. J. JOHNSON. *The Adaptive Decision Maker*. New York: Cambridge University Press, 1993.

PETERSON, R. A. "A Meta-Analysis of Cronbach's Coefficient Alpha." *Journal of Consumer Research* 21 (1994): 381-391.

PRENDERGAST, C. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37 (1999): 7-63.

SHALLEY, C. E. “Effects of Productivity Goals, Creativity Goals, and Personal Direction on Individual Creativity.” *Journal of Applied Psychology* 76 (1991): 179-185.

SHALLEY, C. E. “Effects of Coaction, Expected Evaluation, and Goal Setting on Creativity and Productivity.” *Academy of Management Journal* 38 (1995): 483-503.

SPRINKLE, G. B. “The Effect of Incentive Contracts on Learning and Performance.” *The Accounting Review* 75 (2000): 299-326.

SPRINKLE, G. B. “Perspectives on Experimental Research in Managerial Accounting.” *Accounting, Organizations and Society* 28 (2003): 287-318.

TABLE 1
ANOVA for the Effect of Quantity and Creativity Incentives on Quantity

Panel A: Means (Standard Deviation) for Quantity

	No Creativity Incentive	Creativity Incentive
Quantity Incentive	28.55 (15.13)	17.50 (7.11)
No Quantity Incentive	11.83 (4.25)	11.30 (7.36)

Panel B: Analysis of Variance

Factor	*df*	*Sum of Squares*	*F*	*p-value (two-tailed)*
QUANTITY INCENTIVE	1	2,554.90	28.46	<.01
CREATIVITY INCENTIVE	1	652.74	7.27	.01
QUANTITY × CREATIVITY INCENTIVES	1	538.06	5.99	.02
Error	74			

Panel C: Simple Effects

Effect of CREATIVITY INCENTIVE in *absence* of QUANTITY INCENTIVE	1	2.69	0.03	>.50
Effect of CREATIVITY INCENTIVE in *presence* of QUANTITY INCENTIVE	1	1,221.03	13.60	<.01

TABLE 2
ANOVA for the Effect of Quantity and Creativity Incentives on Average Creativity Ratings

Panel A: Means (Standard Deviation) for Average Creativity Ratings

	No Creativity Incentive	Creativity Incentive
Quantity Incentive	4.51 (0.59)	4.75 (0.45)
No Quantity Incentive	4.91 (0.31)	5.40 (1.03)

Panel B: Analysis of Variance

Factor	*df*	*Sum of Squares*	*F*	*p-value (two-tailed)*
QUANTITY INCENTIVE	1	5.25	12.10	<.01
CREATIVITY INCENTIVE	1	2.57	5.91	.02
QUANTITY × CREATIVITY INCENTIVES	1	0.31	0.71	.40
Error	74			

TABLE 3
ANOVA for the Effect of Quantity and Creativity Incentives on Percentage of Cards Patterned After an Instructional Example

Panel A: Means (Standard Deviation) for Percentage of Cards Patterned After an Instructional Example

	No Creativity Incentive	Creativity Incentive
Quantity Incentive	0.41 (0.21)	0.30 (0.20)
No Quantity Incentive	0.32 (0.18)	0.22 (0.17)

Panel B: Analysis of Variance

Factor	*df*	*Sum of Squares*	*F*	*p-value (two-tailed)*
QUANTITY INCENTIVE	1	0.13	3.46	.07
CREATIVITY INCENTIVE	1	0.23	6.09	.02
QUANTITY × CREATIVITY INCENTIVES	1	0.01	0.01	>.50
Error	74			

TABLE 4
ANOVA for the Effect of Quantity and Creativity Incentives on Weighted Productivity Scores

Panel A: Means (Standard Deviation) for Weighted Productivity Scores

	No Creativity Incentive	Creativity Incentive
Quantity Incentive	122.88 (52.60)	82.87 (31.40)
No Quantity Incentive	57.91 (20.41)	58.56 (35.18)

Panel B: Analysis of Variance

Factor	*df*	*Sum of Squares*	*F*	*p-value (two-tailed)*
QUANTITY INCENTIVE	1	38,782.97	28.17	<.01
CREATIVITY INCENTIVE	1	7,539.83	5.48	.02
QUANTITY × CREATIVITY INCENTIVES	1	8,042.41	5.84	.02
Error	74			

Panel C: Simple Effects

Effect of CREATIVITY INCENTIVE in *absence* of QUANTITY INCENTIVE	1	3.95	0.01	>.50
Effect of CREATIVITY INCENTIVE in *presence* of QUANTITY INCENTIVE	1	16,010.91	11.63	<.01

TABLE 5
ANOVA for the Effect of Quantity and Creativity Incentives on Number of Cards Rated Above 5.5 on Creativity Scale

Panel A: Means (Standard Deviation) of Number of Cards Rated Above 5.5 on Creativity Scale

	No Creativity Incentive	Creativity Incentive
Quantity Incentive	4.30 (2.25)	4.65 (2.56)
No Quantity Incentive	3.39 (1.75)	4.50 (3.14)

Panel B: Analysis of Variance

Factor	*df*	*Sum of Squares*	*F*	*p-value (two-tailed)*
QUANTITY INCENTIVE	1	5.48	0.88	.35
CREATIVITY INCENTIVE	1	10.39	1.67	.20
QUANTITY × CREATIVITY INCENTIVES	1	2.82	0.45	.50
Error	74			

FIGURE 1
Effect of Quantity and Creativity Incentives on Quantity

Mean Quantity:

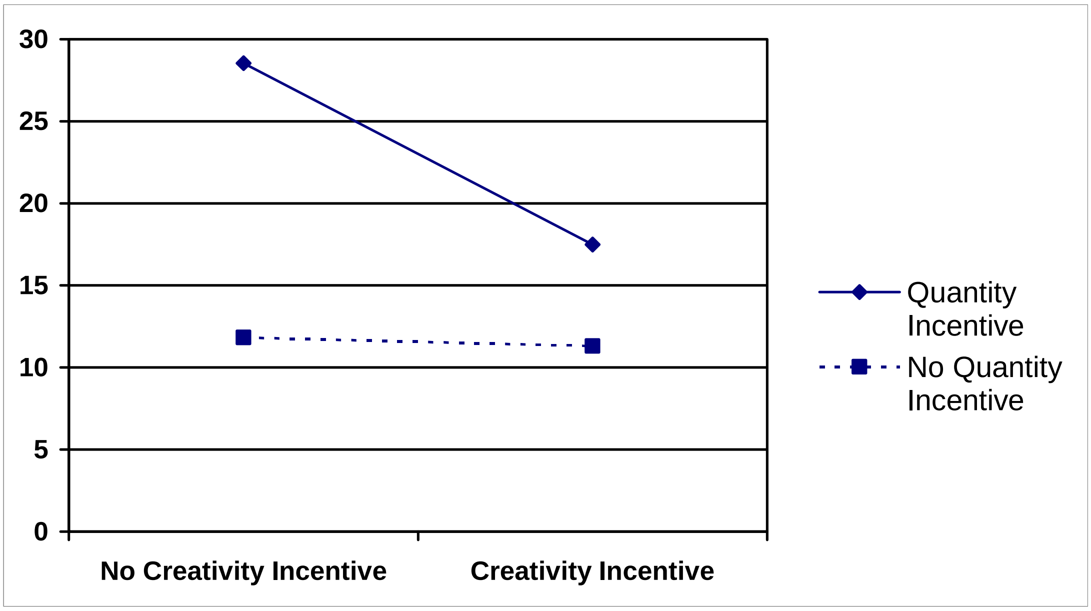


FIGURE 2
Effect of Quantity and Creativity Incentives on Average Creativity Ratings

Mean Average Creativity Rating:

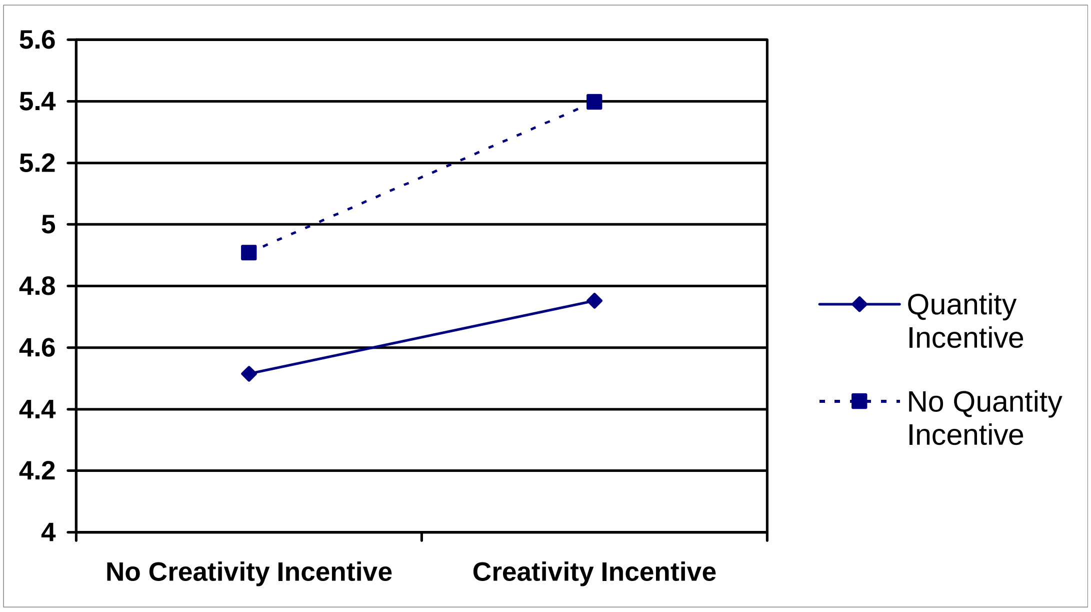


FIGURE 3
Effect of Quantity and Creativity Incentives on Percentage of Puzzles Patterned After an Instructional Example

Means for Percentage of Puzzles Patterned after an Instructional Example:

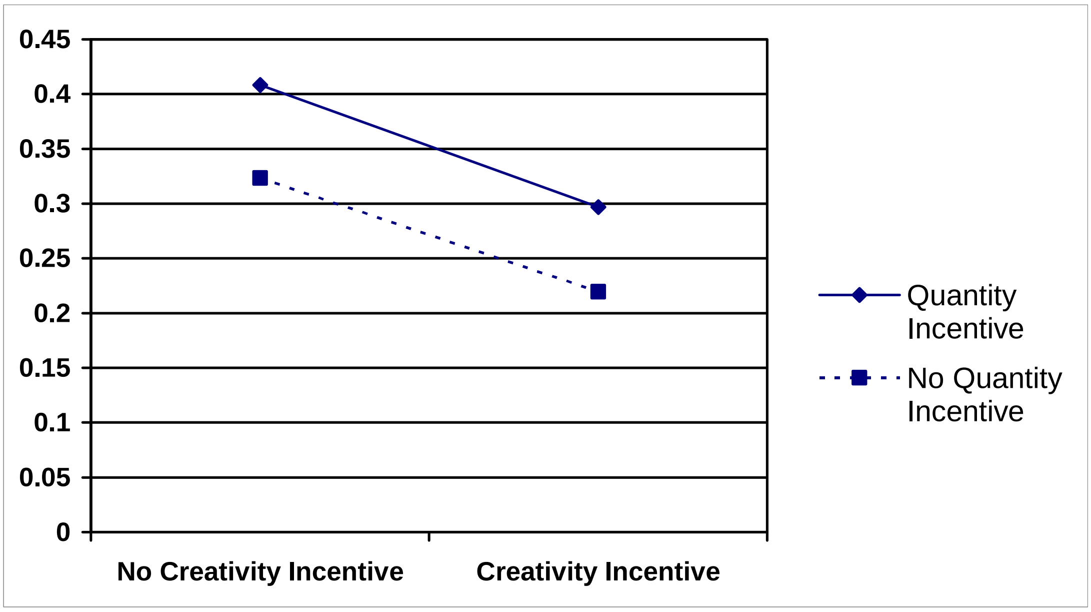


FIGURE 4
Examples of Actual Rebus Puzzles

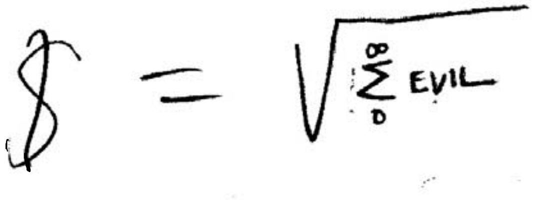
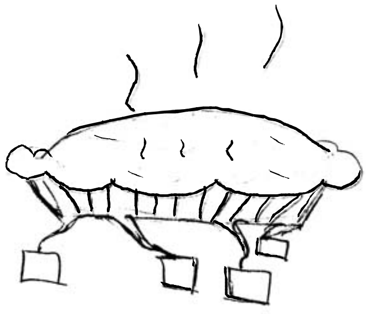
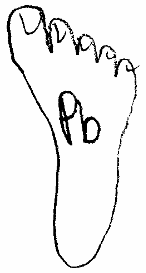
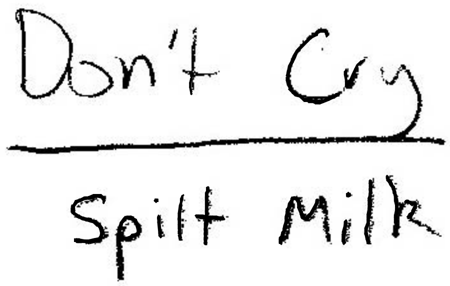
Panel A  **Solution:** Money is the root of all evil **Creativity rating:** 9.18 **Condition:** Creativity-only	**Panel B**  **Solution:** Square root of pi **Creativity rating:** 8.27 **Condition:** Quantity-only
Panel C  **Solution:** Lead foot **Creativity rating:** 7.64 **Condition:** Weighted Quantity × Creativity	**Panel D**  **Solution:** Don't cry over spilt milk **Creativity rating:** 3.91 **Condition:** Quantity-only
Panel E  **Solution:** Chewing gum **Creativity rating:** 1.55 (lowest overall rating) **Condition:** Quantity-only	**Panel F**  **Solution:** Popeye **Creativity rating:** 6.82 **Condition:** Quantity-only (same person who did Panel E)

FIGURE 5

Effect of Creativity and Quantity Incentives on Weighted Productivity Scores

Mean Weighted Productivity:

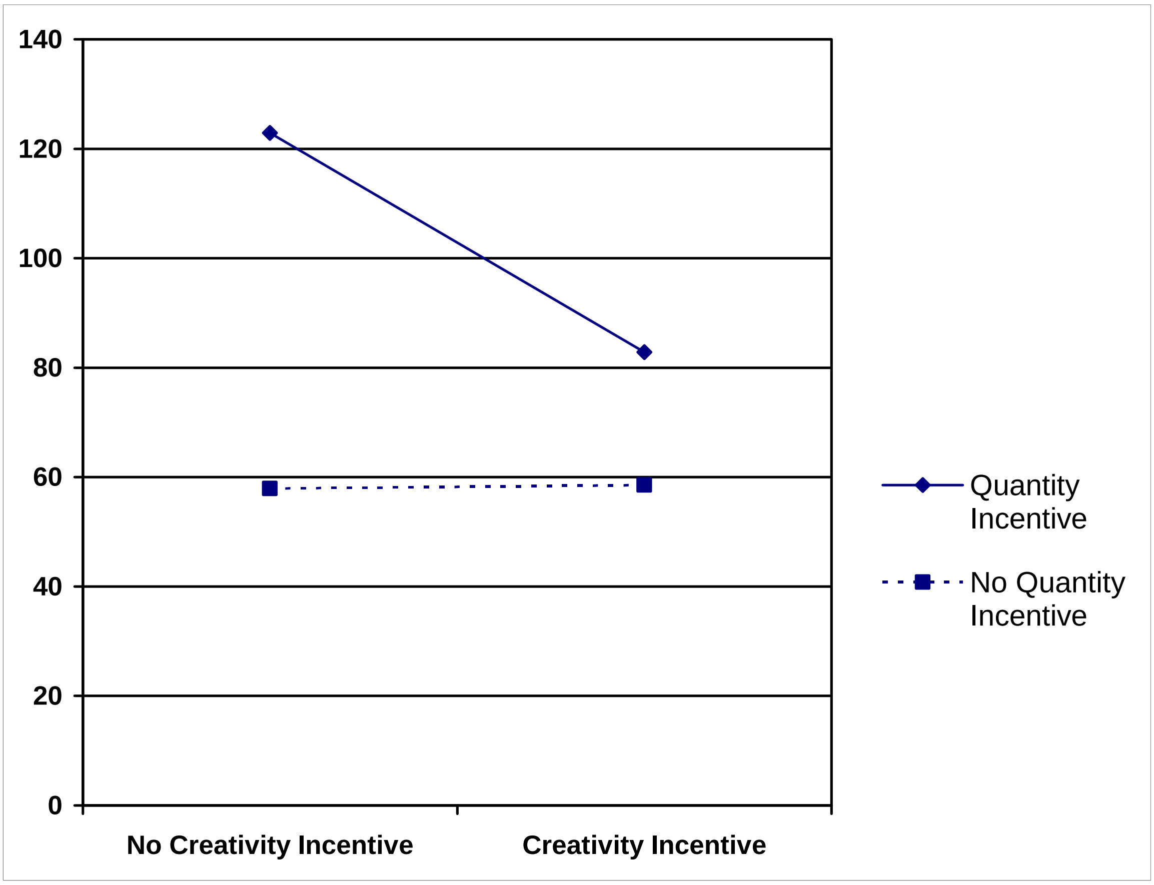
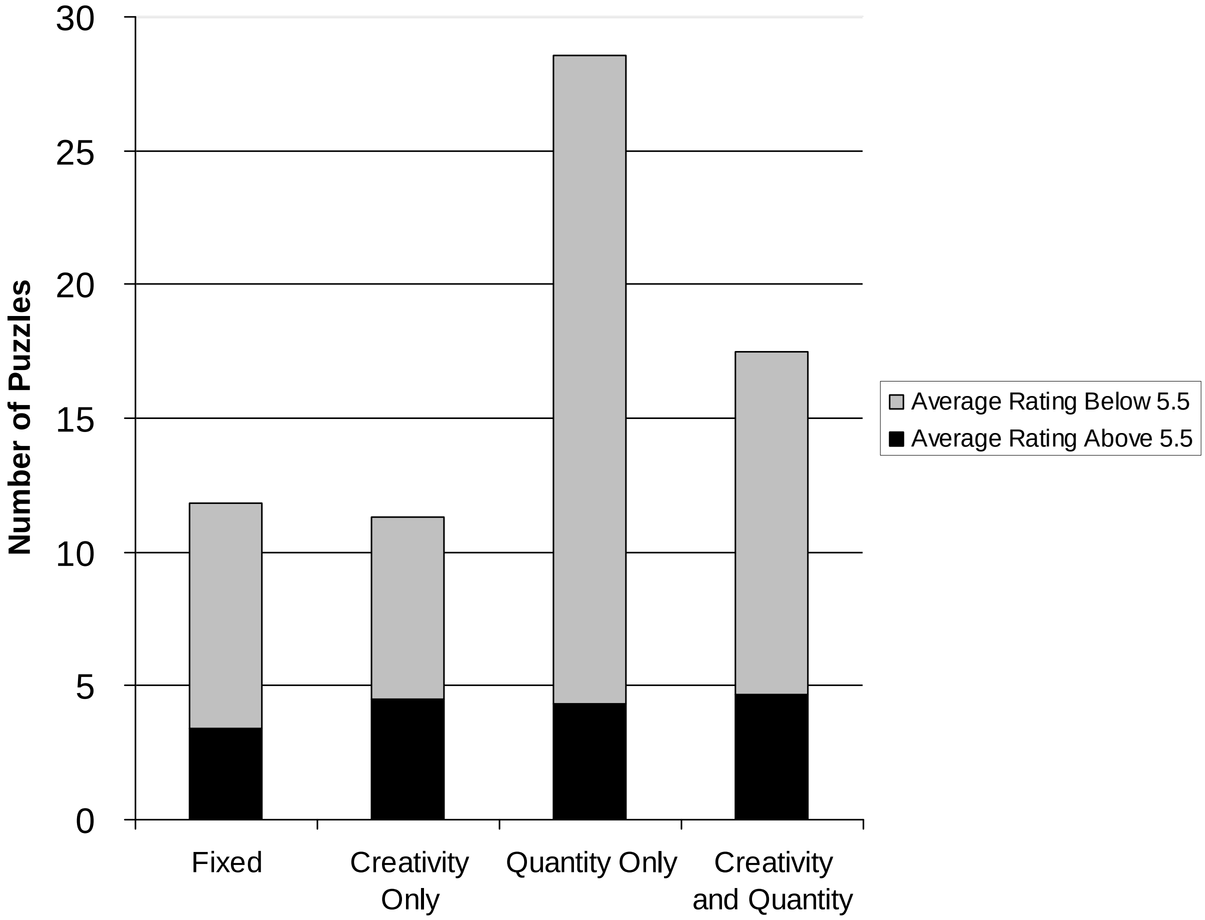


FIGURE 6

Number of Puzzles with a Composite Creativity Rating Above and Below 5.5 by Experimental Condition



This chart illustrates the average number of puzzles each participant produced that received a creativity rating above and below 5.5 in each experimental condition. It highlights that participants compensated for both creativity and quantity produced a comparable number of highly creative puzzles relative to other participants, but produced significantly fewer mediocre puzzles than did participants with quantity incentives only.