# Incentivizing the Creative Process:
# From Initial Quantity to Eventual Creativity

Steven J. Kachelmeier
kach@mail.utexas.edu
The University of Texas at Austin

Laura W. Wang
lauraww@illinois.edu
University of Illinois at Urbana-Champaign

Michael G. Williamson
migwilli@illinois.edu
University of Illinois at Urbana-Champaign

Revised December 2016

# Incentivizing the Creative Process:
# From Initial Quantity to Eventual Creativity

**Abstract:** We conduct a two-stage experimental study to examine the effects of different incentive schemes on high-creativity production. Our primary finding is that, relative to alternative incentive conditions that reward high-creativity production directly, establish a minimum creativity screen, or provide fixed pay, simple quantity-based incentives generate the greatest number of ideas that meet a high-creativity rating threshold. Importantly, this finding *does not* occur at the time of our first-stage experiment when the incentives are in place, but rather arises ten days later at the time of our second-stage request for additional creative ideas. We test alternative process measures to help interpret this result, obtaining the strongest mediation support from a measure of first-stage *divergent thinking*, which we capture from the number of first-stage submissions that differ from the patterns suggested in the instructional examples. Our mediation results suggest that, relative to the other conditions, participants with quantity-based incentives attain an initial advantage in *different* ideas that leads to an eventual advantage in *high-creativity* ideas.

# Incentivizing the Creative Process:
# From Initial Quantity to Eventual Creativity

## I. INTRODUCTION

Management accounting has traditionally viewed performance-based incentives as a direct mapping from what is rewarded to what is desired (e.g., see the review by Bonner, Hastie, Sprinkle, and Young 2000). However, some performance goals require more complex, indirect strategies. A prime example is creativity, which, although generally viewed as a desirable attribute in business (e.g., Fallon and Senn 2006), is not directly responsive to raw effort (Amabile 1996). Put differently, creativity cannot be forced, but rather emerges from a multistage process that involves initial ideas, an "incubation" period, and eventual creative gains (Ward, Smith, and Finke 1999; Csikszentmihalyi and Sawyer 2014). Given this process, it is perhaps not surprising that conventional laboratory experiments have reached mixed findings on whether performance-based incentives improve creative performance (e.g., see review by Byron and Khazanchi 2012). We take a different approach in this study, designing a two-stage experiment to examine whether simple, quantity-based incentives can stimulate the process that eventually leads to gains in high-creativity production, relative to the effects of direct incentives for high creativity, incentives that impose a minimum creativity "screen," or fixed pay.

Our research objective is important because the fruits of business processes, and creative processes in particular, take more time to develop than that available in a typical laboratory experiment. To meet this challenge, our study measures creative production at two points in time: (1) an initial laboratory experiment during which we implement the different incentive conditions and (2) a second-stage request of the same participants ten days later, with no incremental incentive other than $10 fixed pay, which we hold constant across conditions. Our primary finding is that participants with quantity-based pay in the first-stage experiment generate

3

more high-creativity ideas than their counterparts in the other conditions, not at the time of the initial experiment, but ten days later when we elicit their second-stage ideas.

To capture creativity, we adopt the "rebus-puzzle" design task that has been used in previous studies of incentives for creativity (Kachelmeier, Reichert, and Williamson 2008; Kachelmeier and Williamson 2010; Erat and Gneezy 2016a). Like our study, the instructions to these prior 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," providing participants with several illustrative examples. Participants then design their own puzzles, which are later evaluated for creativity by independent raters who are blind to the treatment conditions. The common theme from these prior studies is that, even if quantity or creativity incentives "do no harm" in generating high-creativity ideas, incentives do not significantly *improve* highcreativity production either. We replicate this finding in our first-stage experiment, but reach a different finding ten days later. Specifically, participants with first-stage quantity incentives outperform the *second-stage* high-creativity production of participants with (1) first-stage incentives for high-creativity ideas, (2) first-stage incentives for ideas that achieve a minimally acceptable creativity rating, or (3) fixed pay. To our knowledge, ours is the first study to demonstrate that simple quantity incentives can improve not only total productivity, but also creativity, with the latter effect emerging only after a longer time horizon.

We test the process underlying this finding by establishing that a measure of first-stage *divergent thinking* mediates our second-stage results. We capture divergent thinking by counting the number of first-stage submissions that differ from the themes illustrated in the participants' instructional examples. For example, as one of the instructional examples is the familiar puzzle "man overboard," illustrated by the word "man" written over the word "board," any submitted

4

puzzle with one word written over another word would be consistent with that theme. Divergent thinking requires ideas that depart from the instructional themes to try something new. We find that participants with quantity-based incentives produce more puzzles that meet a divergent-thinking threshold in the first-stage experiment than do participants in any of the other conditions. This initial advantage does not translate to improved first-stage creativity ratings, however, consistent with Runco and Jaeger's (2012) observation that creativity requires ideas that are both original *and* effective.

Ten days later, participants with quantity-based incentives in the first stage turn their initial advantage in divergent thinking into a second-stage advantage in creativity. To support this claim, we find that our first-stage measure of divergent thinking fully mediates the second-stage creativity effect, consistent with the notion that creativity requires an "incubation" period (Csikszentmihalyi and Sawyer 2014). That is, quantity-based participants enjoy a second-stage creativity advantage because they are more successful than other participants at generating new ideas in the first stage. Yet, a raw quantity count by itself *does not* mediate the creativity effect we observe, suggesting that our findings do not simply reflect the inertia of greater initial production volume. Rather, we obtain strong mediation evidence only for a measure of first-stage puzzles that depart from established patterns.

Our findings are related to the gains others have documented from "brainstorming" processes, which can be applied at either the group or individual levels (e.g., Paulus and Yang 2000). The essence of brainstorming is that the best ideas emerge from a process that initially generates a free flow of ideas without restriction or evaluation, providing a foundation for quality gains to be realized later. In our experiment, quantity-based compensation in the first stage provides a monetary incentive for individual brainstorming, including ideas that go beyond the

boundaries of the instructional examples. In contrast, the other compensation schemes we implement either restrict the free flow of ideas (i.e, incentives for high-creativity ideas or incentives for ideas that meet a minimum-creativity screen) or provide no incentive to generate ideas (i.e., fixed pay). One might think that high-creativity incentives would stimulate creative thinking, but the literature has repeatedly shown that attempts to force creativity tend only to generate frustration (Amabile 1996). It is only by incentivizing *all* ideas that we obtain a greater number of divergent ideas in the first-stage experiment that lead to an eventual advantage in second-stage creativity.

Our study contributes to recent progress in strategic applications of performance-based incentives to settings that involve multiple attributes of "good" performance (e.g., Farrell, Kadous, and Towry 2012; Choi, Hecht, and Tayler 2012, 2013). Incentives and controls in such settings must do more than address the end objective. Rather, effective incentives must facilitate the *process* that enables workers to achieve the desired objective. For the goal of high-creativity production, our results indicate that simple quantity-based incentives are best equipped to stimulate the initial divergent ideas that eventually lead to superior creativity.

Section II develops our theory in more detail, Section III explains our research task and manipulations, Section IV presents our primary and mediation results, and Section V concludes.

## II. LITERATURE REVIEW AND THEORY DEVELOPMENT

### Creativity as an Incentivized Goal

Creativity is an elusive goal. Yet, the premise that creative ideas are pivotal to success is a recurring theme in the business literature (e.g., Fairbank and Williams 2001; Fallon and Senn 2006). In accounting, Grabner (2014) reports field evidence that the degree to which firms depend on creativity is positively associated with the use of performance-contingent incentives,

although her study does not identify the specific incentives these firms adopt or their effects on creativity. Thus, although Grabner's (2014) findings suggest that the goal of fostering creativity in business can be *consistent* with the use of performance-based incentives, the question remains open as to exactly what kind of incentives best achieve this goal.

Researchers attempting to establish a causal linkage from incentives to creativity have generally turned to experimental methods. But after decades of research in psychology, management, and more recently accounting and economics, a consensus has yet to emerge (see review by Byron and Khazanchi 2012). Psychologists often advance the argument that creativity cannot be "forced" by incentives, and may in fact be undermined by extrinsic incentives that dampen any intrinsic motivation to be creative (Amabile 1996). The literature supporting this view, however, is generally restricted to consideration of creativity in isolation, rather than the broader notion of creative *production*. An accounting study by Kachelmeier et al. (2008) (hereafter KRW) broadens the question of how incentives affect creativity by examining a laboratory task in which participants design "rebus puzzles." The advantage of this task is meaningful variation in both the quantity of production and the creativity of that production. These dimensions allow KRW to consider different kinds of incentives, including simple quantity-based incentives that reward participants for *all* ideas rather than placing a premium on creative ideas. Relative to creativity-based incentive schemes or fixed pay, KRW find that quantity-based incentives significantly increase the total number of ideas (puzzles) generated by their participants without lowering the number of high-creativity ideas.

While KRW do not find any evidence that quantity-based compensation *harms* high-creativity production, they do not find that quantity-based compensation *improves* high-creativity production either. In a follow-up experiment, Kachelmeier and Williamson (2010) find some

selection effects when participants are allowed to choose their desired pay scheme, but those who choose quantity-based compensation are no less (and no more) creative than other participants over the course of the entire experiment. Most recently, an experiment in the economics literature by Erat and Gneezy (2016a) extends KRW's rebus-puzzle design task to a variety of different creativity-based incentive schemes.[1] They reach the similar conclusion that incentives do not achieve instantaneous improvements in creativity, which they attribute to a "choking hypothesis."

**Creativity as a Process and Quantity-Based Incentives**

We posit that a prime reason for the lackluster incentive effects reported in experiments to date is that creativity requires more time to achieve than that available in a typical laboratory experiment. Even those few experiments that have examined time as a factor have considered differences in minutes, not days. For example, Erat and Gneezy (2016a) find that it does not matter whether they allow participants ten minutes or one hour to complete their experimental task. Any one-shot experiment, however, is unlikely to capture the multistage process that generates creative insight according to psychologists. Building on Hadarnard (1949), Csikszentmihalyi and Sawyer (2014, 79) assert that creative insight first requires an initial "preparation" phase of "vague unease [towards a] diffuse goal," followed by an "incubation" period of seemingly idle time that nevertheless nurtures creativity in the subconscious mind. Similarly, Ward et al. (1999, 191-192) describe creativity as a temporal process that begins with the stimulation of initial, "preinventive structures" and culminates with eventual creative insights. This reasoning suggests that creativity requires two inputs: (1) an initial generation of ideas and (2) the passage of time.

---

[1] Erat and Gneezy (2016a) use the same task and much of the same instructional wording as KRW, adjusted to request only one rebus puzzle per participant rather than several puzzles. See the erratum by Erat and Gneezy (2016b) for acknowledgment of the overlap.

Returning to the effects of incentives, we posit that quantity-based incentives, while not necessarily leading to *initial* gains in creativity, are nevertheless likely to stimulate the initial preparation or generation phase of the creative process. Essentially, quantity-based compensation incentivizes people to produce as much as possible, which in a creative task means thinking of as many ideas as possible. These ideas are unlikely to be particularly creative, but they establish the impetus for eventual creative insights. In commenting on KRW as part of a recent "TedX" video, Matt Diffee, cartoonist for *The New Yorker*, summarizes this notion by stating that the approach he uses to generate creative ideas emphasizes "quantity over quality" (Diffee 2013).[2]

This reasoning is similar to the claimed benefits of brainstorming, which generally holds that the best way to begin a creative process is with an unrestricted generation of multiple ideas. Although brainstorming is generally viewed as a group activity, the principles underlying brainstorming can also be applied at the individual level (Paulus and Yang 2000). In fact, Furnham (2000) argues from his review of the literature dating back to Taylor, Berry, and Block (1958) that individual brainstorming can be *superior* to group brainstorming, insofar as there is less potential for freeriding on others' efforts, intimidation, or "production blocking" that limits output to one group member at a time. Our study does not examine the effectiveness of groups versus individuals, but we do examine the ability of quantity-based incentives to stimulate individual brainstorming by rewarding the generation of as many initial ideas as possible.

Even if effective, quantity-based incentives only establish the first phase of the creative process: idea generation. Our primary contribution beyond related prior experiments is that we also operationalize Csikszentmihalyi and Sawyer's (2014) "incubation" phase of the creative process by waiting ten days until we measure our primary dependent variable. If the creative

---

[2] See the 8:00 minute mark through the 10:00 minute mark of the Diffee (2013) TedX video, available at https://www.youtube.com/watch?v=tAMbxnEtNxE. Although Diffee (2013) somewhat mischaracterizes KRW's findings, his reasoning maps well to the current study.

process works as we reason above, we should see an association between quantity-based incentives and high-creativity production, not at the time the incentives are in place, but several days later after the incubation phase has taken place. Below we consider reasons why alternative incentive schemes might be less likely to generate these same benefits.

**High-Creativity Incentives**

Intuitively, it might seem that the most effective way to generate gains in high-creativity production would be to incentivize high-creativity ideas. This reasoning, however, neglects the psychological reality that creativity is not immediately responsive to raw effort to be creative (Amabile 1996), which in turn is related to Erat and Gneezy's (2016a) "choking" hypothesis. In their original study of incentives to design rebus puzzles, KRW conclude that incentivizing creativity directly leads their participants to limit production to high-creativity efforts, but without generating more high-creativity puzzles than those produced by quantity-compensated participants. Worse, KRW observe that *overall* production suffers under creativity incentives. From a brainstorming perspective, generating fewer initial ideas would be harmful because the brainstorming process works best when initial ideas flow freely without evaluative scrutiny.

**Incentives that Establish a Minimum-Creativity Screen**

As a compromise between rewarding high-creativity ideas only versus rewarding *any* ideas in a quantity-based scheme, an incentive scheme could establish a minimum-creativity screen. In other words, only those initial ideas that meet a minimally acceptable creativity rating would be compensated. Such an incentive would attempt to encourage free thinking while at the same time discouraging "stupid" ideas. While this reasoning has potential merit, it too establishes a restriction on the idea-generation process. That is, any anxiety about meeting the minimum-creativity cutoff could serve to suppress the free flow of ideas. Ultimately, whether

10

any benefits from suppressing clearly bad ideas would outweigh the costs of anxiety about whether ideas would count at the margin is an empirical question that we address in our experimental design.

**Fixed Pay**

A comprehensive study of incentive effects would not be complete without a control condition. We use fixed pay as a control because it holds constant the *presence* of compensation, while removing any performance-based contingency associated with that compensation. For a creative exercise, it is not at all clear *ex ante* that fixed pay would be dysfunctional. A view often advanced in psychology and management is that performance-contingent incentives undermine the intrinsic incentive to be creative (e.g., Amabile 1996; Deci, Koestner, and Ryan 1999). Put differently, if people do a task because it is fun and interesting, the "intrinsic incentives" argument holds that performance-contingent pay can harm performance by making the task seem more like "work." While some studies claim support for this argument (e.g., Deci et al. 1999), this support has not considered the full creative process that we attempt to capture across several days. Whether any gains from maintaining the intrinsic attractiveness of the task under fixed pay would be outweighed by the loss of incentives to generate as many ideas as possible is again an empirical question that our experiment is designed to address.

Overall, we conduct a "horse race" experiment by testing the relative ability of the four incentive schemes discussed above to generate high-creativity production *after* the incentives have been removed, which we capture by asking participants to volunteer any additional creative rebus puzzle ideas that come to mind ten days after the initial, incentivized experiment. Although our reasoning suggests that quantity-based incentives are likely to stimulate the most initial ideas and hence provide the most effective impetus for eventual gains in creativity, we acknowledge

11

that the literature on incentive effects for creative tasks remains too unresolved to support one-tailed tests. Accordingly, we take the conservative approach and test the null hypothesis stated below using nondirectional statistical analyses:

> **Hypothesis** (null form): There will be no difference in high-creativity production from a second-stage request ten days after a first-stage experiment that bases compensation on (1) total first-stage submissions, (2) first-stage submissions that meet a high-creativity threshold, (3) first-stage submissions that meet a minimum-creativity screen, or (4) fixed pay.

**Examining the Creative Process**

Beyond testing our primary hypothesis, we conduct several supplemental analyses to examine the *process* underlying our results. First, if our reasoning is valid that a creativity effect takes more time to establish than that available in a one-shot experiment, we should *not* observe treatment differences in high-creativity production at the time of the first-stage experiment itself. We test our first-stage results to corroborate this reasoning.

Second, we consider alternative process measures that could plausibly mediate our second-stage findings. Specifically, we consider mediators based on (1) first-stage total quantity, (2) a more refined measure of first-stage efforts that indicate divergent thinking, or (3) a measure of task enjoyment. The quantity mediator tests the possibility that generating more initial ideas *in general* can prompt subsequent gains in creativity. The divergent thinking mediator is more nuanced, capturing the reasoning that the seeds of creativity are ideas that differ from the norm. In a sense, divergent thinking provides the "original" part of Runco and Jaeger's (2012) definition of creativity as ideas that are both original and effective. As Runco and Acar (2012, 73) explain, "divergent thinking is not synonymous with creativity," but it establishes a "meaningful potential." Within a multistage creative framework, we examine the extent to which increased divergent thinking stimulates more effective incubation. As explained in more detail

12

later, we measure divergent thinking by counting each participant's first-stage puzzles that differ from the themes illustrated in instructional examples. Finally, we consider the "intrinsic incentives" argument by testing the effects of a self-reported measure of task enjoyment on our findings. To be consistent with our nondirectional primary hypothesis, we test the explanatory power of these alternative process measures as a nondirectional research question:

> **Research question**: To what extent do process measures based on total first-stage quantity, first-stage divergent thinking, or self-reported task enjoyment mediate our second-stage treatment effects?

### III. METHOD, TASK, AND DESIGN

**First-Stage Experimental Task**

Although not the source of our primary dependent variable, we use a first-stage experiment to operationalize initial incentives for a creative task. Participants are 104 volunteers recruited from undergraduate business classes.[3] The task for which participants are compensated in different ways, as detailed below, is to design "rebus puzzles" for 20 minutes. This is the same creative design task developed by KRW (2008) and extended by Kachelmeier and Williamson (2010) and Erat and Gneezy (2016a), hence establishing a common baseline to facilitate interpretation of the second-stage task we describe later. As illustrated in Figure 1 with examples submitted by our participants, a rebus puzzle "is a kind of riddle in which words and/or diagrams are used to represent a familiar term or phrase" (quoted from the experimental instructions).

The rebus puzzle task is creative because participants *design* rebus puzzles rather than solve them, and it also confers the advantage of meaningful variation in the quantity produced. Participants write each rebus puzzle design on a separate index card, indicating the solution to the puzzle at the bottom of the card. Each participant has a stack of blank index cards and an

---

[3] A total of 108 participants completed the first-stage experiment. However, four participants did not return to collect their compensation ten days later, and hence did not complete the second-stage task that generates our primary dependent variable.

"output box" to facilitate this process. To illustrate the task, we provide participants with the same instructional examples of rebus puzzles as those reproduced in the Appendix to KRW (2008, 368-372),[4] informing them that a panel of creativity raters will evaluate their submitted puzzles afterwards for creativity, "where creativity refers to puzzles that are original, innovative, and clever" (quoted from the instructions).

While KRW informed their participants that the researchers value both quantity and creativity, our instructions focus more specifically on *highcreativity production*, informing participants across conditions that "we value the *number of highcreativity puzzles* you can construct" (emphasis in original). The instructions then inform participants (truthfully, based on results reported by KRW) that, "in previous experiments using this task, approximately 15 percent of puzzles received a creativity rating at or above 6," such that "a rating at or above 6 would be considered a high-creativity puzzle." This statement provides all participants with the same objective to submit as many highcreativity puzzles as possible. However, their *economic incentive* to submit such puzzles differs by condition, as explained next.

**First-Stage Experimental Design**

Our research design manipulates the nature of participants' *first-stage* compensation for submitting rebus puzzles. Although participants are informed of their compensation rule at the time of the first-stage experiment, they are instructed to collect their compensation ten days later at a specified location. In addition to giving us time to determine quantity counts and creativity ratings as needed to determine compensation (see below), this structure provides an opportunity to access the same participants for a second-stage experimental request. As described later, this second request comprises our primary dependent variable, although participants are unaware of the second-stage request at the time of the first-stage experiment.

---

[4] As noted by KRW, these examples come from public sources such as Morris (1983) and Griggs (2000).

*Quantity-Based Incentives*

To generate as many ideas as possible, our first experimental treatment rewards the *raw quantity* of rebus puzzles submitted in the first-stage experiment. Specifically, participants in this condition are informed that we will "*count the number of rebus puzzles you submit, no matter what creativity ratings those puzzles receive*" (quoted from instructions, with emphasis in original). The instructions further inform participants that we will determine a cash payment rate to result in $45 for the participant submitting the most rebus puzzles in this condition and $5 for the participant submitting the least, thus generating "an expected average compensation around $25." Similar to KRW, a payment structure of this nature allows us to hold constant the *amount* of average compensation across conditions, while varying the *nature* of that compensation.

*High-Creativity Incentives*

To test the ability of incentives to motivate high-creativity production directly, a "highcreativity" treatment condition informs participants that they will be compensated only for puzzles that receive a composite creativity rating of six or higher (on a ten-point scale) by an independent panel. Thus, like those in the quantity-based condition, participants in the high-creativity condition receive more money for more puzzles, but for these participants, only high-creativity submissions "count." This condition is new to the current study, as the creativity-based compensation schemes used in prior research reward creativity in different ways, such as the "average creativity" and "creativity-weighted" conditions in KRW.[5] In contrast to these approaches, the puzzles submitted in our high-creativity condition either count fully for compensation, if rated at six or higher, or are not rewarded at all. As in the quantity-based condition, participants in the high-creativity condition are informed that the participant

_____

[5] The "creativity-weighted" condition in KRW rewards all puzzles submitted, but assigns compensation weights based on creativity ratings.

15

submitting the most (least) puzzles rated six or higher will receive $45 ($5), thus generating average compensation of approximately $25.

### *Minimum-Creativity Screen*

As communicated to participants in the experimental instructions, prior research indicates that rebus puzzle designs achieve a creativity rating of six or higher only approximately 15 percent of the time. Thus, the high-creativity incentive condition imposes a relatively stringent compensation screen. To test the effects of a more forgiving screen, we operationalize a "minimum-creativity" condition that is identical to the structure of the high-creativity condition except that it rewards all puzzles rated *four* or higher. As explained in Section II, the intent of this scheme is to evaluate whether an effective compromise is possible that encourages idea generation but guards against the potential for exploiting a quantity-based scheme by submitting a large number of low-creativity puzzles. As with the other performance-based compensation schemes, the minimum-creativity condition establishes a linear payment formula that generates a payment of $45 ($5) for the participant with the most (least) puzzles rated four or higher, thus again generating average compensation of around $25.

### *Fixed Pay*

Finally, as an experimental control, we implement a fixed-pay condition in which all participants receive $25 simply for participating. Thus, like their counterparts with performance-based pay, the fixed-pay participants receive compensation at the same average amount ($25), and read the same instructional note that the experimenters "value the number of high-creativity puzzles you can construct." However, fixed-pay participants have no economic incentive to maximize either quantity or creativity. Although neither quantity nor creativity measures are necessary to determine compensation in this condition, the instructions inform the fixed-pay

participants that they too will have to come to a different location ten days later to receive their fixed payment because "different versions of the research require waiting, and we want to pay all participants at the same time" (quoted from the instructions).

We conducted the first-stage experiment in a controlled setting on the same day, with random assignment of the materials corresponding to each of the four incentive conditions. Participants were informed that different participants had different instructions, although participants were unaware of the nature of these differences. A pre-experimental question asked participants to explain briefly in writing how they would be compensated, and an experimenter visited each individual station to verify accurate responses as a manipulation check.[6] We determined compensation in the manner indicated in the instructions, treating each condition as a separate compensation pool under that condition's rules. As explained above, participants in each condition earned average first-stage compensation of approximately $25, by construction.

**Second-Stage Experiment**

Although we implement our treatment manipulation in the first-stage experiment described above, we obtain our primary dependent variable ten days later when participants return to collect their compensation. We provide a window of several hours at a specified location for participants to do so. Upon arrival, an experimenter provides the participant with a packet containing an instructional document that begins as follows:

> Thank you again for participating in our experiment last week. You will receive your cash compensation for last week's session in just a few minutes. However, if you are willing to provide responses to just a few more items, we will pay you an additional $10 in cash today, on top of what you have already earned. We expect that these additional responses will require no more than 15 minutes.

---

[6] In the few instances of incomplete or inaccurate responses, an experimenter wrote down the missing information and obtained a visual confirmation of the correct understanding.

The instructions proceed with a request for at least one and up to ten additional creative rebus puzzles,[7] along with a post-experimental questionnaire. The packet includes eleven blank index cards for participants to use for their additional rebus puzzles. The additional $10 fixed pay, held constant across conditions, serves as a goodwill gesture to compensate willing participants for our additional request. Participants received this additional $10 irrespective of the number of second-stage puzzles submitted, which participants placed in a large envelope along with any unused index cards. No participant turned down the offer of $10 in exchange for completing our additional request, which is understandable given that a participant could comply with our request by submitting as little as one additional puzzle in addition to completing the questionnaire. After handing in their second-stage envelopes, participants were paid their first-stage compensation plus $10 for the second-stage task, as promised.

Two points are important to emphasize about the second-stage administration. First, participants completed the second-stage task before learning the amount of their first-stage compensation, with the exception that fixed-pay participants knew they would be paid $25. This procedure helps to minimize the influence of reactions to first-stage compensation outcomes on second-stage performance. Second, participants had no prior knowledge of the second-stage request for additional rebus puzzles until that request was made. This design choice achieves closer theoretical alignment with the "incubation" phase of the creative process (Csikszentmihalyi and Sawyer 2014). The alternative of notifying participants at the time of the first-stage experiment that we would ask them to design additional rebus puzzles ten days later could have led participants to conjecture (perhaps incorrectly) about our research objectives and any long-term incentives to be creative. Moreover, participants could have been tempted to

_____

[7] More specifically, participants are asked to submit one "most creative" puzzle, along with up to ten additional puzzles, at the participant's discretion. Separating the puzzles designated by participants as being their "most creative" efforts does not yield any incremental treatment effects beyond those reported.

"cheat" in the ensuing days by searching for rebus puzzles that would not have been the participants' own creations. While long-term creativity incentives are certainly worthy of study, our research focuses on the creative incubation process that follows participants' *initial* incentives. To maintain this focus, we do *not* implement different incentive structures in the second-stage, but instead offer the same $10 additional fixed stipend to all participants in exchange for their willingness to complete the second-stage request.

**Creativity Ratings**

To construct our dependent variable, we obtain creativity ratings from four panels of eight raters each – three panels for the first-stage experiment and a fourth panel for the second-stage puzzles. Creativity raters are student volunteers from business honors classes that did not provide experimental participants. We pay each rater $50 for a rating session of about 2½ hours. Creativity raters first read the same background instructions and examples as those provided to the experimental participants, but without any information on treatment manipulations. Raters then use individual radio-frequency response devices to rate each puzzle, projected one at a time, on a 1 (lowest) to 10 (highest) scale. Four panels are needed to mitigate rater fatigue, as each panel evaluates approximately 500 rebus puzzles. We randomize the order of puzzles for rating purposes, and raters are blind to treatment conditions. To ensure similar calibration across panels, each rating session begins with the same set of 40 puzzles. For the fourth session of 240 secondstage puzzles, we add 170 first-stage puzzles (randomly selected and interspersed) to maintain a similar distribution of puzzles across rating sessions.

Because creativity ratings are subjective and noisy, we drop the highest and lowest individual ratings for each puzzle to reduce variability. Thus, each puzzle's rating reflects the average rating awarded to that puzzle by six of the eight panelists.

# IV. RESULTS

## First-Stage Results

Although our hypothesis development focuses on the second-stage task, we begin with an analysis of the first-stage experiment to establish a baseline that is comparable to prior research. Table 1 tallies and Figure 2 depicts means by condition for first-stage total production and high-creativity production, with the latter defined as the number of puzzles with a composite rating of six or higher. As is apparent in Figure 2, participants with quantity-based pay dominate the other conditions in the total number of first-stage puzzles submitted, with an average of 24.15 vs. 13.56 across the other three conditions. This finding is not surprising, given that the quantity-based compensation condition rewards total quantity, and is consistent with the findings of prior experiments.[8] Also consistent with prior research, the advantage in total quantity achieved by participants with quantity-based pay does not extend to *high-creativity* production. As Table 1 and Figure 2 indicate, the average of 2.35 first-stage high-creativity puzzles submitted is similar across conditions, with the possible exception of only 1.56 first-stage high-creativity puzzles submitted, on average, in the condition with a minimum-creativity screen.

ANOVAs confirm the observations summarized above. The ANOVA on first-stage total quantity in Table 2, Panel A indicates an overall treatment effect ($F_{3,100} = 12.56$; $p < 0.01$), with the follow-up test in Panel B confirming that total quantity is higher in the quantity-based pay condition than in the other three conditions ($F_{1,100} = 37.28$; $p < 0.01$),[9] with the other three conditions not significantly differing from each other ($F_{2,100} = 0.20$; $p = 0.82$). In contrast, the ANOVA on first-stage *high-creativity* production in Table 2, Panel C does not reach statistical

---

[8] The finding that quantity-based pay results in the greatest number of rebus puzzles submitted is common to KRW, Kachelmeier and Williamson (2010), and the current study. This finding does not apply to Erat and Gneezy (2016a) because they asked each participant to submit only one rebus puzzle.

[9] Throughout, all reported *p*-values are two-tailed.

significance at conventional levels ($F_{3,100} = 1.71$; $p = 0.17$). As the means in Table 1 indicate, the greatest number of first-stage high-creativity puzzles occurs in the condition with fixed pay (average of 2.73), followed by participants with high-creativity incentives (2.65), quantity-based pay (2.44), and a minimum-creativity screen (1.56). Given the lack of statistical significance, we do not comment further on these differences. Overall, we conclude that quantity-based pay leads to more first-stage quantity but not more first-stage creativity than the other conditions. As noted next, we reach a different conclusion in the second stage.

**Second-Stage Results**

Figure 3 depicts the means summarized in Table 1 for second-stage total quantity and high-creativity production. To focus on creativity that emerges in the second stage, we first delete 10 puzzles (4.2 percent of the total) that are duplicates of the same participants' first-stage puzzles, although our statistical conclusions are not sensitive to this action. For total quantity, the overall second-stage average of 2.30 puzzles per participant is a small fraction of the first-stage production. This difference is to be expected, given the different nature of the two sessions. That is, the first-stage task is conducted as a formal experiment with 20 minutes dedicated to designing rebus puzzles. The time participants devote to our second-stage request, in contrast, is purely at their discretion, with no advance notice. That is, participants could complete the second-stage request with as little as one new rebus puzzle, and many did just that. Nevertheless, given that that this same environment applies across conditions, we are able to conduct statistical comparisons.

As in the first stage, participants with first-stage quantity-based pay submit more puzzles in the second stage (average of 2.93) than do participants in the other conditions, although participants with first-stage high-creativity incentives come close (average of 2.77). Participants

in the minimum-creativity and fixed pay conditions lag behind with second-stage total quantity averages of 1.48 and 2.00, respectively. However, the ANOVA for second-stage quantity in Table 3, Panel A does not reach statistical significance for an overall treatment effect ($F_{3,100} = 1.59$; $p = 0.20$).[10] Thus, at least statistically, it appears that any effect of first-stage incentives on second-stage total quantity is modest.

Our hypothesis, however, is directed to second-stage *high-creativity* production. Here, the advantage of participants with first-stage quantity-based pay is more dramatic and statistically significant, with 1.48 average second-stage high-creativity puzzles in the quantity-based pay condition vs. an average of 0.71 across the other three conditions. In percentage terms, slightly over half of the second-stage puzzles submitted by participants with first-stage quantity incentives meet the high-creativity threshold of a rating of six or higher (i.e., 1.48/2.93, or 51 percent), as compared to 35 percent in the condition with first-stage high-creativity incentives (i.e., 0.96/2.77), 33 percent in the condition with first-stage minimum-creativity incentives (i.e., 0.48/1.48), and 35 percent in the condition with fixed pay (i.e., 0.69/2.00).

The ANOVA in Table 3, Panel B confirms an overall treatment difference in second-stage high-creativity production at a level that is at least borderline significant ($F_{3,100} = 2.53$; $p = 0.06$). We obtain a more significant difference when comparing the quantity-based pay condition to a composite of the other three conditions in Panel C ($F_{1,100} = 5.96$; $p = 0.02$). The other three conditions do not significantly differ from each other ($F_{2,100} = 1.20$; $p = 0.30$), such that we combine them in our supplemental mediation analyses reported below. Overall, our findings for second-stage high-creativity production are in stark contrast to our first-stage results,

---

[10] Although the omnibus *F*-statistic for total second-stage quantity is not statistically significant, because the quantity-based condition has the highest mean, we test this condition against the other three conditions to verify that the other conditions are not the sole reason for the lack of significance. The quantity-based vs. other three conditions comparison for second-stage quantity also fails to reach statistical significance $F_{1,100} = 1.96$; $p = 0.16$.

in which the quantity-based pay condition does not even achieve the highest mean, much less a statistically significant advantage. Thus, consistent with the reasoning that creativity requires a free flow of initial ideas followed by an "incubation" phase, we reject the null hypothesis in favor of the conclusion that first-stage quantity incentives do in fact lead to greater creativity, not at the time of the first-stage experiment, but only long after the incentives have been removed. As such, relative to the other incentive conditions we test, quantity-based pay appears to be the most effective at motivating the *process* that leads to eventual creativity. We next turn to our supplemental research question to explore this process in more detail by considering alternative candidates for mediating this result.

**Mediation Analyses**

*Quantity-Based Mediation*

We consider alternative process variables that could plausibly mediate and hence help to explain the finding that participants with first-stage quantity-based incentives achieve greater second-stage high-creativity production than do participants in the other conditions. The first and seemingly most obvious mediation candidate is quantity itself. That is, if quantity-based pay motivates eventual creativity simply because it generates the greatest number of first-stage ideas, a first-stage quantity measure should mediate our second-stage findings. To test this possibility, we first conduct a path analysis to follow the traditional mediation approach suggested by Baron and Kenny (1986). Results are summarized in Figure 4, Panel A. Although the path analysis indicates a statistically significant association between first-stage quantity-based pay and first-stage quantity ($p < 0.01$), as reported earlier, the relation between first-stage quantity and second-stage high-creativity production is not statistically significant at conventional levels ($p = 0.13$), thus failing the second step of the mediation process. We also conduct a bootstrapping-based

23

mediation analysis as recommended by MacKinnon, Lockwood, Hoffman, West, and Sheets (2002) and illustrated by Preacher and Hayes (2004; 2008). The bootstrapping-based test across 5,000 resamples does not support a significant mediating effect of total quantity, insofar as the 95 percent confidence interval includes zero. We conclude that our primary result is unlikely to be explained solely as the inertia of first-stage quantity leading to second-stage creativity. As reported next, however, we obtain much stronger mediation results using a refined measure based on the quantity of first-stage puzzles that evidence divergent thinking.

### *Divergent-Thinking Mediation*

As explained in our theoretical development, creativity requires ideas that are both original (i.e., different) and effective (i.e., appealing) (Runco and Jaeger 2012). Divergent thinking generates ideas that are original, but not necessarily effective. Accordingly, as Runco and Acar (2012) explain, divergent thinking is not synonymous with creativity, but it is a necessary precursor to creativity. This reasoning suggests that a measure of first-stage divergent thinking, meaning initial ideas that differ from the norm, could be a more effective mediator for the creative process than mere quantity, which counts *all* ideas.

As a proxy for divergent thinking, we count the number of each participant's first-stage puzzle submissions that are *not* patterned after one of the eleven puzzles that we adopt from KRW as instructional examples. That is, given the availability of eleven instructional examples, a straightforward approach to the design task would be to extend the instructional examples along similar lines. For instance, the submitted puzzle "too close for comfort" in Panel C of Figure 1 is a simple extension of the example "too funny for words" in the instructions, as both puzzles use the homonyms "two" and "four" to represent "too" and "for," respectively, writing out the first word in the puzzle two times and the second word four times. We establish specific

24

criteria to identify extensions for all instructional examples.[11] Submissions *not* meeting these criteria qualify as "divergent thinking" puzzles, which we tally individually for each participant. For example, the submission "house call" in Panel C of Figure 1, while not viewed by our raters as being particularly creative (composite rating of 4.0), nevertheless qualifies as a divergent thinking puzzle because it is unlike any of the instructional examples.

Table 1 reports and Figure 5 depicts the average number of divergent thinking puzzles submitted in the *first-stage* experiment, by condition. As Figure 5 indicates, participants in the quantity-based pay condition generate significantly more puzzles that qualify under our proxy for divergent thinking than do those in any of the other conditions, even though this advantage in divergent thinking does not translate to an advantage in high-creativity production at the time of the first-stage experiment. First-stage divergent thinking does, however, mediate our primary finding that quantity-based pay leads to improved *second-stage* creativity. As documented in the path analysis of Figure 4, Panel B, the first requirement of mediation is satisfied because quantity-based pay is significantly positively associated with first-stage divergent thinking, relative to the other three conditions ($p < 0.01$). The second requirement of mediation is also met, as first-stage divergent thinking is significantly positively associated with *second-stage* high-creativity production ($p = 0.03$). Finally, the third requirement of mediation is satisfied, as the treatment effect of quantity-based pay on second-stage high-creativity production is no longer statistically significant in a combined model that includes the treatment factor as well as the

---

[11] Specifically, we classify submissions as extensions of the instructional examples if they meet any of the following criteria, with the terms in quotation marks appearing in the solution (original puzzle solutions in parentheses): (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 or puns based on academic "degrees" (three degrees below zero); (6) uses of the numbers two and four to represent the words "to," "too," or "for" (too funny for words); (7) words with letters "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 (hole in one); or (11) something crossed out using the "not" symbol (to be or not to be). See the Appendix to KRW (2008, 369-370) for the instructional examples common to their study and ours, which in turn are adapted from public sources such as Morris (1983) and Griggs (2000).

mediator variable for first-stage divergent thinking ($p = 0.25$), suggesting full mediation. A bootstrapping-based mediation test (Preacher and Hayes 2004; 2008) based on 5,000 resamples supports mediation based on a 95-percent confidence interval that does not include zero.

Our conclusion from the mediation analyses to this point is as follows. The use of quantity-based pay in our first-stage experiment motivates greater quantity in general. But more importantly, quantity-based pay motivates participants to venture further beyond the finite number of patterns suggested by the instructional examples. These new ideas are not necessarily attractive enough to be judged highly creative at first. Ten days later, however, participants with quantity-based pay are able to translate their initial advantage in divergent thinking into a subsequent advantage in creativity. We do not mean to imply that participants *literally* transform their first-stage divergent puzzles into second-stage creative puzzles, but the evidence does suggest that increased first-stage divergent thinking expands the potential for creative incubation. As such, the ability of quantity-based incentives to stimulate an eventual creativity effect does not appear to reflect the inertia of sheer volume *per se* as much as it reflects an increased willingness to think outside the norm. As further support for this conclusion, in contrast to the results for our measure of divergent thinking, we find no evidence of mediation from the number of first-stage puzzles that do *not* meet our criteria for divergent thinking. As shown in Figure 4, Panel C, the association of such submissions with second-stage high-creativity production is near zero and far from statistically significant ($p = 0.82$), while the effect of quantity-based pay remains highly significant in a combined model ($p = 0.02$).

*Enjoyment-Based Mediation*

Finally, given the focus in psychology and management on the tradeoff between intrinsic and extrinsic motivation (e.g., Amabile 1996; Deci et al. 1999), we evaluate how a self-reported

measure of task enjoyment bears on our findings. Specifically, participants respond to the seven-point Likert-scale question "I enjoyed creating rebus puzzles" immediately after the first-stage experiment. The means of 5.41, 5.19, 5.49, and 6.00 in the quantity-based pay, high-creativity pay, minimum-creativity screen, and fixed-pay conditions, respectively, do not statistically differ in an omnibus test across all four conditions ($F_{3,99}$ = 1.77; $p$ = 0.16, untabulated).[12] Indeed, the quantity-based pay condition does not even have the highest mean, thus failing the first test of mediation. Nevertheless, for completeness, we test a combined model with second-stage high-creativity production as the dependent variable that includes the treatment factors and the self-reported task-enjoyment measure as explanatory variables. Results (untabulated) indicate that the compensation treatment effect remains statistically significant in favor of quantity-based pay ($F_{1,101}$ = 6.42; $p$ = 0.01). We also observe a modest positive association between self-reported first-stage task enjoyment and second-stage high-creativity production, albeit at a level that only approaches statistical significance ($F_{1,101}$ = 2.62; p = 0.11). We conclude that the advantage participants with quantity-based pay achieve in second-stage high-creativity production occurs *despite* any effect of task enjoyment, not because of it.

Accordingly, while the effect of task enjoyment on creativity is interesting in its own right, enjoyment does not appear to explain our finding that first-stage quantity-based pay maximizes second-stage high-creativity production. The overall conclusion to our supplemental research question is that, of the process measures we consider, a measure of first-stage divergent thinking provides the most convincing evidence of mediation.[13] Put differently, participants with

---

[12] We also elicit self-reported task enjoyment after the *second-stage* task, reaching the same conclusion from an omnibus test that the four treatment conditions do not significantly differ ($F_{3,99}$ = 1.50; $p$ = 0.21), although the fixed-pay condition nominally has the highest mean. For both the first-stage and second-stage comparisons, we control for an *ex ante* measure of anticipated task enjoyment from a pre-experimental questionnaire as a covariate. This covariate does not statistically differ across conditions ($F_{3,100}$ = 0.21; $p$ = 0.74).

[13] In addition to these process measures, the second-stage post-experimental questionnaire also elicits a self-reported measure of time spent thinking about the rebus-puzzle task after the first-stage experiment. Means are 34 minutes,

quantity-based pay appear to generate more high-creativity efforts in the second stage because they are more willing to experiment with different ideas in the first stage.

**Other Supplemental Analyses**

*Testing the Condition × Stage Interaction*

Although we report first-stage results as a baseline for comparison, our hypothesis and primary analyses are directed to second-stage high-creativity production. As reported above, we find that first-stage quantity-based pay is associated with increased high-creativity production in the second-stage task, but not in the first-stage experiment. By themselves, however, these tests do not necessarily provide statistical evidence that the first and second stages differ from each other. Accordingly, in this subsection we test a composite model that combines the first and second stage data to test for a condition × stage interaction. If our reasoning is correct that first-stage quantity incentives motivate the *process* that eventually leads to an advantage in creativity, we should see evidence of this process by directly comparing the two experimental stages.

A challenge in combining the first- and second-stage data is that the two stages are not directly comparable. As previously noted, the first-stage task is a formal experiment in which participants generate an average of 16.2 rebus puzzles, whereas the second-stage task is a shorter, voluntary exercise in which participants submit an average of only 2.3 puzzles. Accordingly, we standardize observations within each stage by applying a Z-score transformation of high-creativity production, subtracting the overall stage mean and dividing by the standard deviation. This transformation results in an overall mean of zero within each stage,

---

19 minutes, 30 minutes, and 27 minutes in the quantity-based incentives, high-creativity incentives, minimum-creativity screen, and fixed-pay conditions, respectively. Although participants with quantity-based pay self-report spending the most time thinking about the task, which is consistent with our primary finding, the difference between the quantity-based condition and a composite of the other three conditions is not statistically significant ($F_{1,102} = 2.04$; $p = 0.16$), even after using a rank transform to lessen the effects of outliers. Thus, self-reported time is of limited value to explain our treatment effect, which likely reflects the considerable noise in this measure.

removing the stage main effect by construction. Nevertheless, we can still test the compensation scheme × stage *interaction* to examine whether quantity-based pay exerts a different treatment effect across the two experimental stages.

Results from a repeated-measures analysis of standardized high-creativity production (untabulated) indicate that the effect of quantity-based pay vs. a composite of the other three conditions interacts with the experimental stage ($F_{1,102} = 3.31$; $p = 0.07$). While the (twotailed) statistical significance of this interaction is at a borderline level, when combined with the findings reported earlier, the evidence points to a different effect of first-stage quantity incentives on first- versus second-stage high-creativity production. Indeed, the quantity-based condition does not even generate the highest mean in the first-stage experiment (see Table 1), whereas the same condition is significantly greater than the other three conditions for the second-stage task. We conclude that quantity-based incentives need time to incubate before they generate improvement in high-creativity production relative to the other incentive conditions we test.

### *Alternative Cutoff for High Creativity*

We define "high-creativity" puzzles as those rated six or higher because prior research finds that only about 15 percent of puzzle submissions achieve such a rating, which is also what the instructions communicate to participants. That being said, six is barely above the midpoint of the ten-point creativity scale used by our raters, such that it is worthwhile to corroborate our findings using a more stringent high-creativity cutoff. Accordingly, we repeat our primary analyses after redefining "high-creativity production" as puzzles reaching a composite rating of *seven* or higher. Results (untabulated) confirm that first-stage quantity-based pay is associated with more second-stage high-creativity puzzles than in the other three conditions ($F_{1,100} = 6.37$; $p = 0.01$), whereas the other conditions do not statistically differ from each other ($F_{2,100} = 1.75$;

*p* = 0.18). We also obtain the same conclusions from our mediation analyses. In sum, even though puzzles rated seven or higher are very rare,[14] we find that our results are robust to this alternative cutoff.[15]

### *Alternative Measure of High-Creativity Production*

Our primary measure of high-creativity production reflects both a high creativity rating (i.e., puzzles rated six or higher) and a production measure (i.e., the number of such puzzles submitted). However, the nature of our second-stage task limits the inferences one can draw from a continuous measure of production. That is, our second-stage request only asks for at least one additional puzzle, and indeed, the modal participant submitted exactly one puzzle. Accordingly, our primary measure of high-creativity production is limited by the discrete nature of the dependent variable. To address this concern, we classify each participant as submitting (1) zero second-stage puzzles that meet the high-creativity threshold, (2) one high-creativity puzzle, or (3) two or more high-creativity puzzles. This classification enables us to apply logistic regression to the categorical frequency counts in lieu of the conventional ANOVA-based approach.

Figure 6 shows the percentage frequencies for high-creativity productivity by cell. As this figure indicates, participants with first-stage quantity-based incentives have the highest percentage frequency of two or more second-stage high-creativity submissions (i.e., 30 percent, as opposed to an average of 12 percent across the other three cells). In addition, the quantity-based condition evidences the *lowest* frequency of zero second-stage high-creativity submissions (i.e., 33 percent, as opposed to an average of 52 percent across the other three cells). An ordered

---

[14] Second-stage means by condition are 0.74. 0.46, 0.12, and 0.42 puzzles rated seven or higher in the quantity-based pay, high-creativity incentives, minimum-creativity incentives, and fixed-pay conditions, respectively.
[15] In contrast, our second-stage findings are not robust to defining "high-creativity" puzzles as those rated *five* or higher, as the advantage of the first-stage quantity-based condition over the other three conditions is not statistically significant at conventional levels ($F_{1, 102} = 2.32$; *p* = 0.13). Accordingly, the incentive effect we observe is robust to a more stringent cutoff for high-creativity production, but not to a less-stringent cutoff.

logistical regression (untabulated) of the categorical frequencies in Figure 6 confirms that participants with first-stage quantity-based pay achieve greater second-stage high-creativity production than do participants in the other three conditions (z = 2.16; $p$ = 0.03). Thus, our primary findings appear to be robust to a more categorical definition of high-creativity production.

## V. CONCLUSIONS

The management accounting literature on performance-based incentives has evolved over the years from examining the direct effects of incentives on a routine production task with one dimension that responds well to raw effort (e.g., Chow 1983) to the more strategic effects of incentives on complex production tasks with multiple dimensions that respond to effort in different and subtle ways (e.g., Farrell et al. 2012; Choi et al. 2012, 2013; KRW 2008). We continue this progress by examining the effects of initial incentives on the process that generates creative ideas, using the "rebus puzzle" design task originated by KRW. In particular, the primary advantage of our approach is that we examine high-creativity production at two points in time: (1) a first-stage experiment during which the different incentive conditions are implemented, and (2) a second-stage request of the same participants for any additional ideas they are willing to volunteer ten days later, long after the incentives have been removed. This approach allows us to peer closer inside the creative *process* that requires both stimulation and incubation to succeed (Ward et al. 1999; Csikszentmihalyi and Sawyer 2014).

Our primary finding is that simple first-stage quantity-based incentives generate increased second-stage high-creativity production relative to alternative conditions with high-creativity incentives, a minimum creativity screen, or fixed pay. Evidencing the creative process, this difference occurs only for our second-stage request, not at the time of the first-stage

31

experiment. Supplemental process-based analyses indicate that our primary findings are not mediated by simple first-stage quantity counts or a measure of task enjoyment. Rather, we obtain the strongest mediation support from a measure of first-stage *divergent thinking*, capturing the number of first-stage ideas that differ from the instructional examples even if they are not sufficiently appealing to meet the high-creativity threshold at the time. In simple terms, creativity requires ideas that are both different (i.e., original) and appealing (i.e., effective) (Runco and Jaeger 2012). Relative to the other conditions, quantity-based pay motivates a greater initial willingness to generate *different* ideas, but only ten days later do we observe an advantage among these participants in the generation of ideas that meet the high-creativity rating threshold.

Our experiment is subject to limitations. Most prominently, although we examine high-creativity production at two points in time, we chose not to examine long-term incentives *per se*. That is, the incentive conditions we implement apply *only* at the time of our first-stage experiment, not at the time of our second-stage request that provides our primary dependent variable. Indeed, participants are not even aware during our first-stage experiment of our intent to request additional creative ideas ten days later. This design choice helps us to control for possible alternative interpretations, insofar as an explicit long-term incentive scheme might have tempted our participants to exert effort outside the creative process, such as by searching online sources. As such, our study attempts to capture the creative process that emerges naturally after an initial stimulation, as opposed to a process that is governed more explicitly by long-term incentives. Still, we acknowledge that long-term incentives are worthy of examination, providing a fruitful avenue for future research.

We see the primary contribution of our research as helping to resolve conflicting views on the effectiveness of monetary incentives when tasks are defined by "soft" quality

characteristics such as creativity. Specifically, although a widely held view in psychology is that explicit incentives undermine intrinsic creativity (e.g., Amabile 1996), in accounting, Grabner (2014) has recently reported field evidence that creativity-intensive firms are *more* likely than other firms to utilize performance-based incentive schemes. Our study suggests that a possible reconciliation of these seemingly opposing claims is that incentives can take time to be effective. Consistent with this theme, we encourage further exploration of incentive effects over longer horizons than those typically afforded in a one-shot laboratory experiment.

**FIGURE 1**
**Examples of Rebus Puzzles Submitted by Participants**
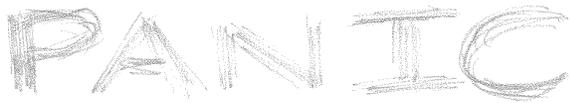
**Panel A: High-Creativity Puzzles**
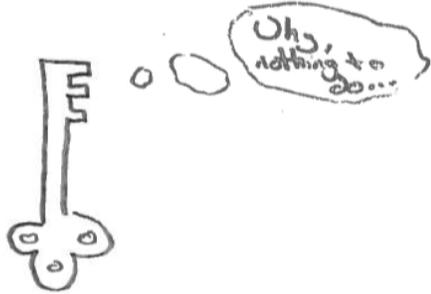

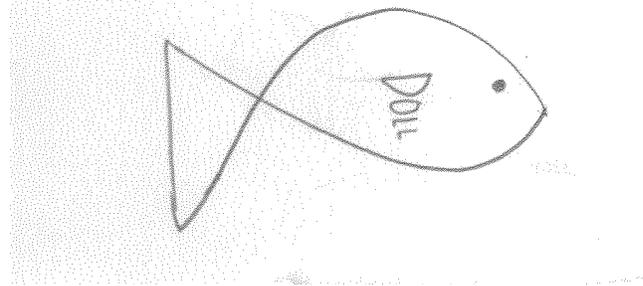
**Solution:** Keyboard
**Creativity rating:** 8.17
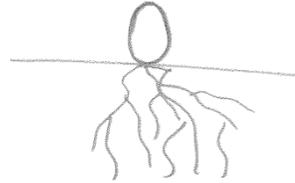


**Solution:** Eggplant
**Creativity rating:** 7.83
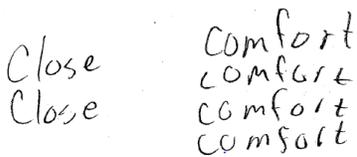
**Panel B: Borderline High-Creativity Puzzles**
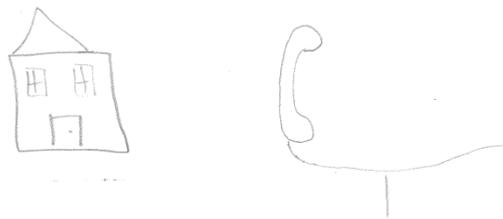


**Solution:** Widespread panic
**Creativity rating:** 6.00



**Solution:** Dolphin
**Creativity rating:** 6.00

**Panel C: Moderate-Creativity Puzzles**
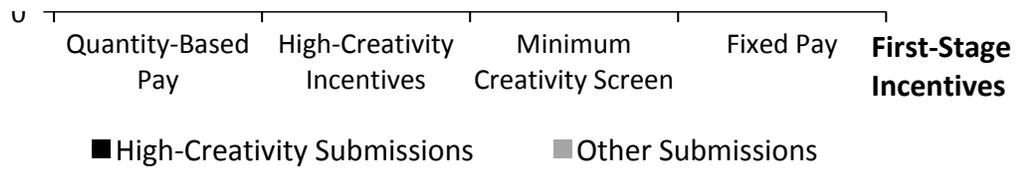


**Solution:** Too close for comfort
**Creativity rating:** 4.67



**Solution:** House Call
**Creativity rating:** 4.00

**FIGURE 2**
**First-Stage Production by Condition**

**First-Stag**

0 ─┬─────────────┬─────────────┬─────────────┬─────────────┬─

| Quantity-Based Pay | High-Creativity Incentives | Minimum Creativity Screen | Fixed Pay | **First-Stage Incentives** |

■ High-Creativity Submissions    ■ Other Submissions

This figure graphs first-stage high-creativity submissions and other submissions by condition, with "high-creativity" submissions defined as puzzles that receive a composite creativity rating of six or higher.
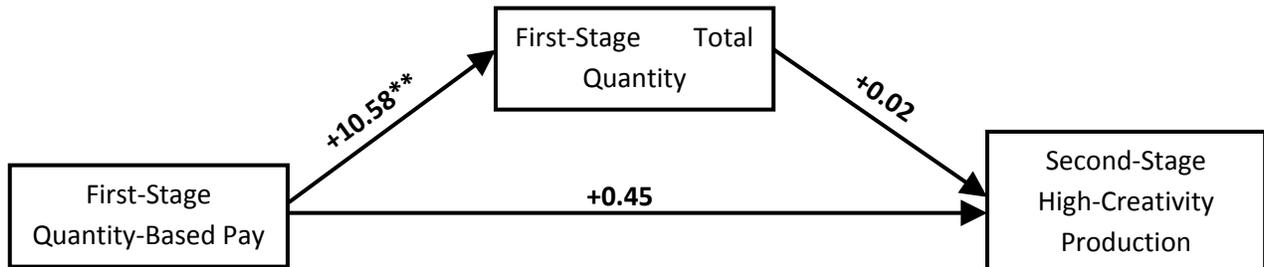
**FIGURE 3**
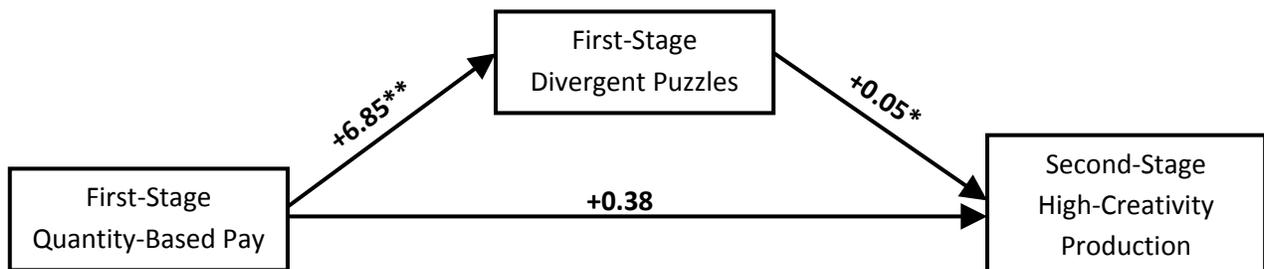**Second-Stage Production by Condition**
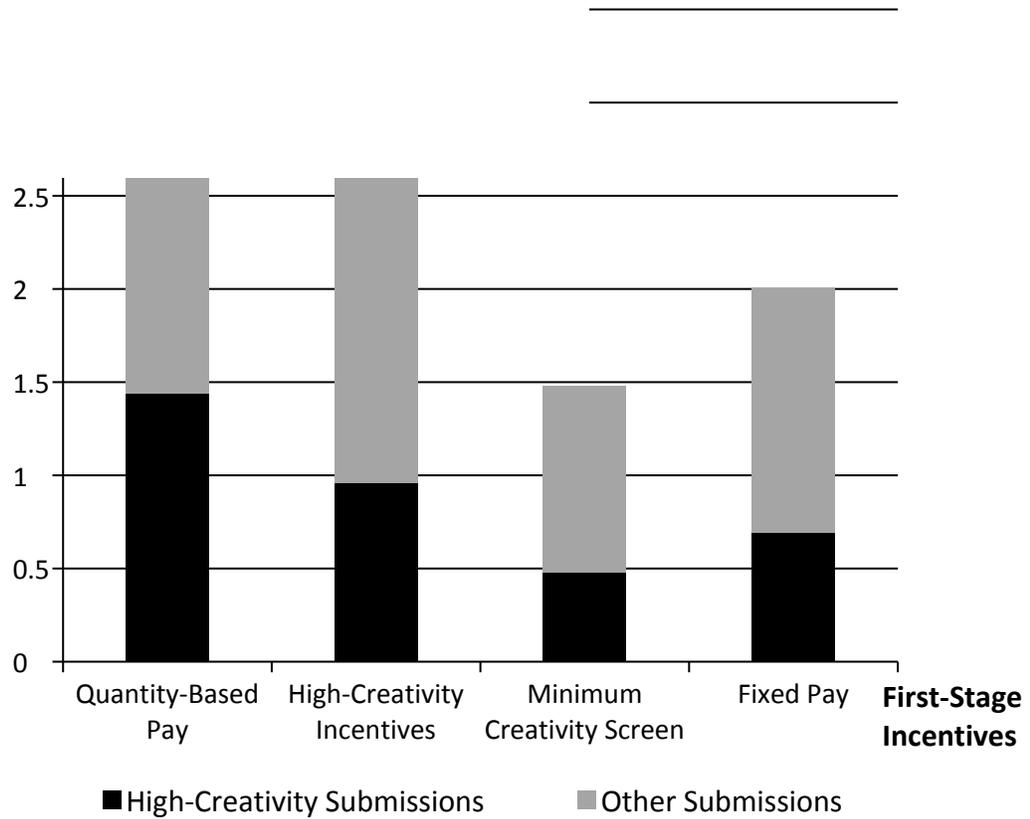


This figure graphs second-stage high-creativity submissions and other submissions by condition, with "high-creativity" submissions defined as puzzles that receive a composite creativity rating of six or higher.

# FIGURE 4
## Mediation Path Analyses for First-Stage Quantity and Divergent Thinking
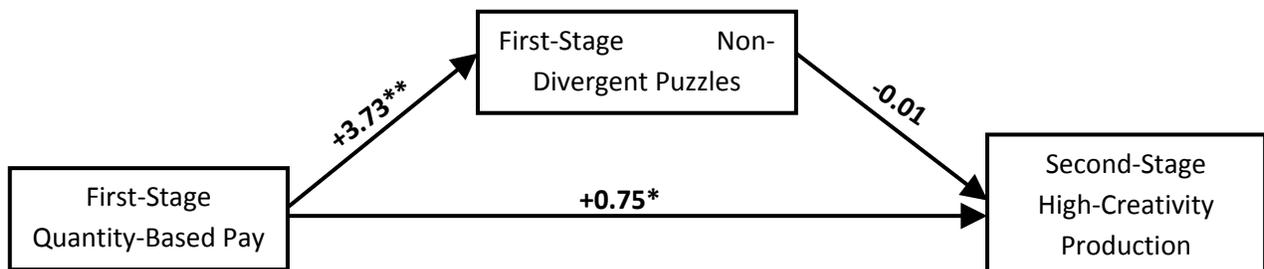
**Panel A: Path Analysis for the Mediation Effect of First-Stage Total Quantity**



**Panel B: Path Analysis for the Mediation Effect of First-Stage Divergent-Thinking Puzzles**



**Panel C: Path Analysis for the Mediation Effect of First-Stage Non-Divergent Puzzles**



This figure shows path coefficients from a combined model that regresses second-stage high-creativity production on a 0, 1 indicator for quantity-based pay (coded 0 across the other three conditions) and the proposed mediator. In turn, the mediator is the total number of first-stage puzzles submitted in Panel A, the number of first-stage puzzles that evidence divergent thinking (i.e., departures from the instructional patterns) in Panel B, and the number of first-stage puzzles that do *not* evidence divergent thinking in Panel C.

* and ** indicate statistical significance at the 0.05 and 0.01 levels, respectively.

**FIGURE 5**
**First-Stage Divergent Thinking by Condition**

**First-Sta**



| | | | | First-Stage Incentives |

■ Divergent-Thinking Submissions    ▨ Other Submissions
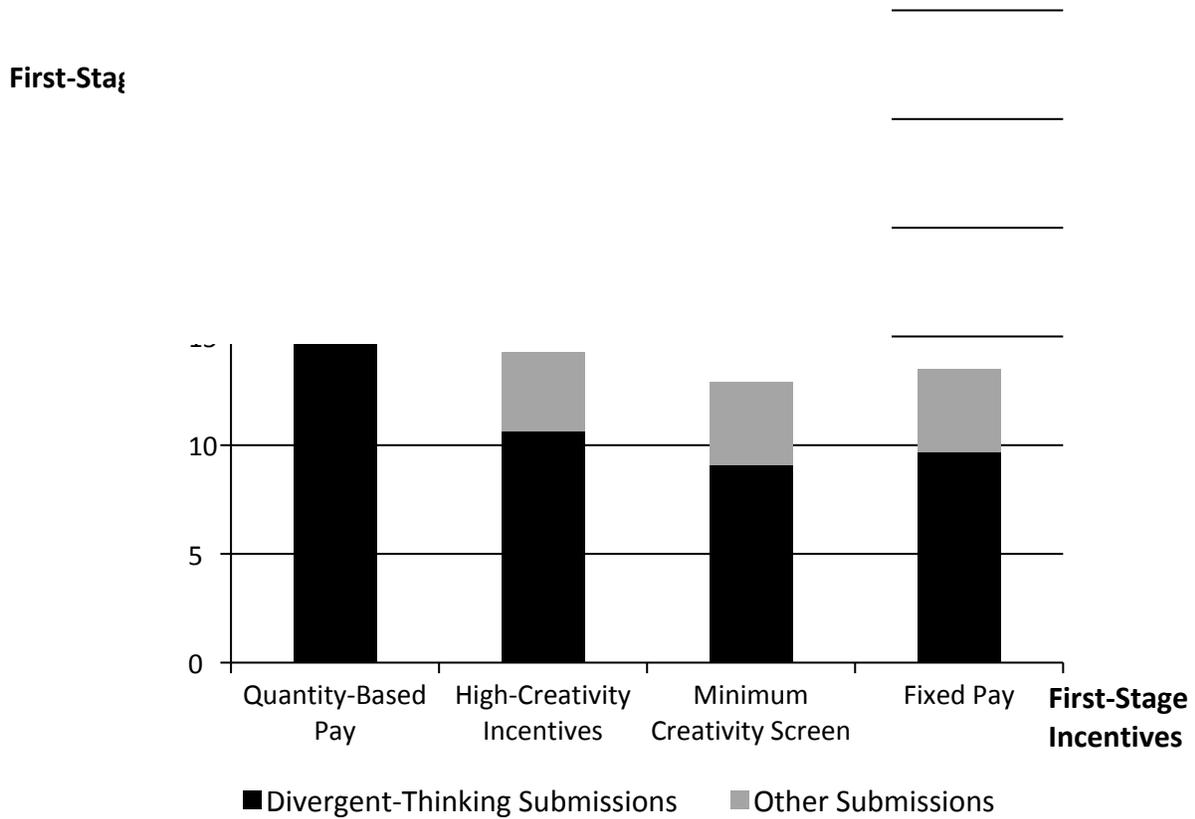
This figure graphs first-stage divergent-thinking submissions and other submissions by condition, with "divergent-thinking" submissions defined as puzzles that do not follow one of the patterns illustrated in the instructional examples.

**FIGURE 6**
**Categorical Percentage Frequencies of Second-Stage High-Creativity Production**
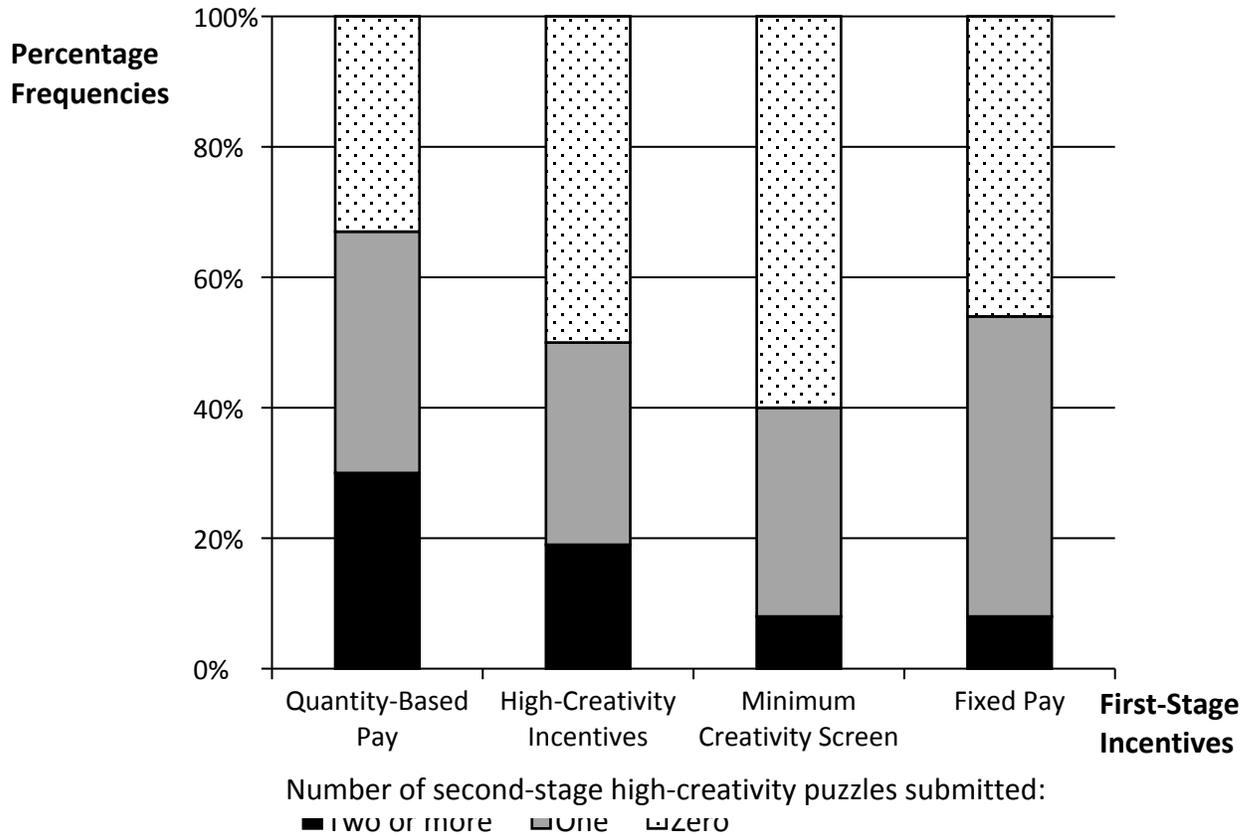


This figure shows the cumulative percentage frequencies within each treatment condition for which participants submitted two or more second-stage high-creativity puzzles (black), one highcreativity puzzle (gray), or no high-creativity puzzles (dot pattern).

**TABLE 1**

**Means and (Standard Deviations) by Condition**

| | **First-Stage Incentives** | | | | |
| | **Quantity-Based Pay** | **High-Creativity Incentives** | **Minimum Creativity Screen** | **Fixed Pay** | **Overall Averages** |
|---|---|---|---|---|---|
| First-stage total quantity submitted | 24.15 (12.21) | 14.27 (5.97) | 12.92 (5.11) | 13.50 (4.93) | 16.21 |
| First-stage high-creativity production[a] | 2.44 (2.37) | 2.65 (2.33) | 1.56 (1.73) | 2.73 (1.73) | 2.35 |
| First-stage divergent-thinking puzzles[b] | 16.67 (8.72) | 10.62 (4.75) | 9.12 (4.00) | 9.69 (4.01) | 11.53 |
| Second-stage total quantity submitted | 2.93 (2.83) | 2.77 (3.14) | 1.48 (0.96) | 2.00 (2.28) | 2.30 |
| Second-stage high-creativity production[a] | 1.44 (1.85) | 0.96 (1.56) | 0.48 (0.65) | 0.69 (0.88) | 0.89 |
| N[c] | 27 | 26 | 25 | 26 | |

---

[a] High-creativity production is the number of puzzles submitted that receive a creativity rating of six or higher.
[b] Divergent-thinking puzzles are submissions that do not exhibit one of the patterns illustrated in the instructional examples.
[c] Reported cell sizes reflect the number of participants who completed both the first and second stages of the experiment. Four participants, comprising one in the high-creativity incentives condition, two in the minimum-creativity screen condition, and one in the fixed-pay condition, completed the first-stage experiment but did not return for the second stage. Hence, our primary dependent variable is unavailable for these four participants, such that we exclude them from all reported analyses.

**TABLE 2**

**Effects of First-Stage Compensation on First-Stage Production**

**Panel A: Omnibus ANOVA for First-Stage Total Quantity**

| Source of variance: | df | MSE | *F* | *p*-value[b] |
|---|---|---|---|---|
| Incentive conditions[a] | 3 | 753.22 | 12.56 | < 0.01 |
| Error | 100 | 59.99 | | |

**Panel B: Follow-up Comparisons for First-Stage Total Quantity**

| | df | MSE | *F* | *p*-value[b] |
|---|---|---|---|---|
| Quantity-based pay vs. the other three conditions | 1 | 2236.26 | 37.28 | < 0.01 |
| Comparison across the other three conditions | 2 | 11.70 | 0.20 | 0.82 |

**Panel C: Omnibus ANOVA for First-Stage High-Creativity Production[c]**

| Source of variance: | df | MSE | *F* | *p*-value[b] |
|---|---|---|---|---|
| Incentive conditions[a] | 3 | 7.34 | 1.71 | 0.17 |
| Error | 100 | 4.30 | | |

---

[a] Omnibus ANOVA comparisons are across four treatment conditions that are manipulated in the first-stage experiment: quantity-based pay, high-creativity incentives, a minimum creativity screen, and fixed pay.
[b] All reported *p*-values are nondirectional (i.e., two-tailed).
[c] High-creativity production is the number of puzzles submitted that receive a creativity rating of six or higher.

**TABLE 3**

**TABLE 3**

**Effects of First-Stage Compensation on Second-Stage Production**

**Panel A: Omnibus ANOVA for Second-Stage Total Quantity**

| Source of variance: | df | MSE | *F* | *p*-value[b] |
|---|---|---|---|---|
| Incentive conditions[a] | 3 | 9.78 | 1.59 | 0.20 |
| Error | 100 | 6.14 | | |

**Panel B: Omnibus ANOVA for Second-Stage High-Creativity Production[c]**

| Source of variance: | df | MSE | *F* | *p*-value[b] |
|---|---|---|---|---|
| Incentive conditions[a] | 3 | 4.54 | 2.53 | 0.06 |
| Error | 100 | | | |

**Panel C: Follow-up Comparisons for Second-Stage High-Creativity Production[c]**

| | df | MSE | *F* | *p*-value[b] |
|---|---|---|---|---|
| Quantity-based pay vs. the other three conditions | 1 | 10.66 | 5.96 | 0.02 |
| Comparison across the other three conditions | 2 | 1.49 | 1.20 | 0.30 |

---

[a] Omnibus ANOVA comparisons are across four treatment conditions that are manipulated in the first-stage experiment: quantity-based pay, high-creativity incentives, a minimum creativity screen, and fixed pay.
[b] All reported *p*-values are nondirectional (i.e., two-tailed).
[c] High-creativity production is the number of puzzles submitted that receive a creativity rating of six or higher.

## References

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

Baron, R. M., and D. A. Kenny. 1986. Moderator-mediator variables distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 512 (6): 1173-1182.

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.

Byron, K., and S. Khazanchi. 2012. Rewards and creative performance: A meta-analytic test of theoretically derived hypotheses. *Psychological Bulletin* 138 (4): 809-830.

Choi, J., G. W. Hecht, and W. B. Tayler. 2012. Lost in translation: The effects of incentive compensation on strategy surrogation. *The Accounting Review* 87 (4): 1135-1163.

Choi, J., G. W. Hecht, and W. B. Tayler. 2013. Strategy selection, surrogation, and strategic performance measurement systems. *Journal of Accounting Research* 51 (1): 105-133.

Chow, C. W. 1983. The effects of job standard tightness and compensation scheme on performance: An exploration of linkages. *The Accounting Review* 58 (4): 667-685.

Csikszentmihalyi, M., and K. Sawyer. 2014. Creative insight: The social dimension of a solitary moment. In *The Systems Model of Creativity*, edited by M. Csikszentmihalyi, 73-98. Dordrecht, The Netherlands: Springer Science and Business Media.

Deci, E. L., R. Koestner, and R. M. Ryan. 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125 (6): 627-668.

Diffee, M. 2013. How to get a great idea: Matt Diffee at TEDxRedding. Video. Available online at https://www.youtube.com/watch?v=tAMbxnEtNxE.

Erat, S., and U. Gneezy. 2016a. Incentives for creativity. *Experimental Economics* 19 (2): 269280.

Erat, S., and U. Gneezy. 2016b. Erratum to: Incentives for creativity. *Experimental Economics*, forthcoming. Available at http://link.springer.com/article/10.1007/s10683-016-9495-y.

Fairbank, J. F., and S. D. Williams. 2001. Motivating creativity and enhancing innovation through employee suggestion system technology. *Creativity and Innovation Management* 10 (2): 68-74.

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. K. Kadous, and K. L. Towry. 2012. Does the communication of causal linkages improve effort allocations? An experimental investigation. *Journal of Management Accounting Research* 24: 77-102.

Furnham, A. 2000. The brainstorming myth. *Business Strategy Review* 11 (4): 21-28.

Grabner, I. 2014. Incentive system design in creativity-dependent firms. *The Accounting Review* 89 (September): 1729-1750.

Griggs, R. A. 2000. A one-minute 'intelligence' test. *Teaching of Psychology* 27: 132-134.

Hadarnard, J. 1949. *The Psychology of Invention in the Mathematical Field*. Princeton: Princeton University Press.

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.

Kachelmeier, S. J., and M. G. Williamson. 2010. Attracting creativity: The initial and aggregate effects of contract selection on creativity-weighted productivity. *The Accounting Review* 85 (September): 1669-1691.

MacKinnon, D. P., C. M. Lockwood, J. M. Hoffman, S. G. West, and V. Sheets. 2002. A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods* 7: 83-104.

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

Paulus, P. B., and H-C Yang. 2000. Idea generation in groups: A basis for creativity in organizations. *Organizational Behavior and Human Decision Processes* 82 (1): 76-87.

Preacher, K. J., and A. F. Hayes. 2004. SPSS and SAS procedures for estimating indirect effect in simple mediation models. *Behavior Research Methods, Instruments & Computers* 36: 717731.

Preacher, K. J., and A. F. Hayes. 2008. Contemporary approaches to assessing mediation in communication research. In *The Sage Sourcebook of Advanced Data Analysis Methods for Communication Research*, edited by A. F. Hayes, M. D. Slater, and L. B. Snyder, 13-54. Thousand Oaks, CA: Sage Publications.

Runco, M. A., and S. Acar. 2012. Divergent thinking as an indicator of creative potential. *Creativity Research Journal* 24 (1): 66-75.

Runco, M. A., and G. J. Jaeger. 2012. The standard definition of creativity. *Creativity Research Journal* 24 (1): 92-96.

Taylor, D. W., P. C. Berry, and C. H. Block. 1958. Does group participation when using brainstorming facilitate or inhibit creative thinking? *Administrative Science Quarterly* 3 (1): 2347.

Ward, T. B., S. M. Smith, and R. A. Finke. 1999. Creative cognition. In *Handbook of Creativity*, edited by R. Sternberg, 189-212. New York, NY: Cambridge University Press.