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

Editor's note: Accepted by Kathryn Kadous under the Senior Editorship of Mark L. DeFond
Submitted October 2015
Accepted June 2018

We appreciate comments on previous versions from Editor Kathryn Kadous, two anonymous reviewers, E.B. Altiero, Scott Asay, Spencer Anderson, Tim Bauer, Jason Brown, Will Ciconte, Dan Collins, Angelo Ditillo, Stephanie Grant, Andy Kitto, Lisa LaViers, Justin Leiby, Michael Majerczyk, Bill Messier, Sharon Noppe, Andreas Ostermaier, Kathy Rupar, Tom Vance, Brian White, and Xinyu Zhang. We also received helpful comments from participants at the 2015 Management Accounting Section Midyear Meeting, the 2016 American Accounting Association Annual Meeting, and workshop participants at Bocconi University, the University of Florida, Georgia State University, the University of Graz, Harvard University, the University of Illinois at Urbana-Champaign, Indiana University, the University of Iowa, Maastricht University, the University of Nevada at Las Vegas, the University of Notre Dame, the Ohio State University, Technical University of Munich, the University of Texas at Arlington, and the University of Texas at Austin. The first author is grateful for the support of the Randal B. McDonald Chair in Accounting, and the third author is grateful for the support of the A. C. Littleton Professorship in Accountancy.

Incentivizing the Creative Process:
From Initial Quantity to Eventual Creativity

Abstract: In two experiments, we test the potential for performance-contingent incentives to facilitate the creative *process* by enhancing the initial preparation that precedes creative incubation. The defining characteristic of both experiments is a second-stage task that is separated in time from the first-stage implementation of different incentive schemes. In Experiment 1, the second stage takes place ten days after we implement conditions with quantity-based pay, high-creativity incentives, incentives that establish a minimum-creativity threshold, and a control condition with fixed pay. In Experiment 2, we test the effects of incentives with a much shorter incubation period of only 20 minutes, during which an experimenter escorts participants on a relaxing walk in between compensated work periods. In both experiments, we find that participants with quantity incentives outperform the high-creativity production of their fixed-pay counterparts only in the second-stage task, after the incubation period. Mediation analyses suggest that quantity-incentivized participants' propensity to try more divergent ideas in the first stage sparks their creativity advantage in the second stage.

Keywords: creative process, incentives, incubation, long-term production

JEL Classifications: D24, D91, M11, M41

Incentivizing the Creative Process:
From Initial Quantity to Eventual Creativity

I. INTRODUCTION

Despite several experiments in psychology, organizational behavior, and more recently, management accounting and economics, the question of how performance-based incentives affect creativity remains controversial (see review by Byron and Khazanchi 2012). The mixed evidence to date likely reflects the fact that creativity is not easily forced. Rather, creativity involves a multistage process encompassing initial preparation, detached incubation, and eventual creative gains (Wallas 1926; Armbruster 1989; Csikszentmihalyi and Sawyer 2014). Accordingly, we direct two experiments to the potential for performance-based incentives to improve high-creativity production by facilitating the creative process.[1] Both experiments employ a two-stage design in which quantity incentives "prime the pump" by prompting a greater number of first-stage ideas that lead to second-stage gains in creativity. Although we do not observe any positive effects of incentives on high-creativity production at the time the incentives are initially applied, incentivized participants generate more high-creativity ideas than participants with fixed pay when we bring them back for a follow-up task ten days later in Experiment 1, or just 20 minutes later after a relaxing walk in Experiment 2.

Our research objective is important because contemporary business demands workers who are not only productive on routine tasks, but who can also generate high-creativity ideas (e.g., Fallon and Senn 2006; Speckbacher 2017). Because creativity does not necessarily respond to raw effort (Amabile 1996), there is reason to believe that incentives cannot *directly* improve creativity. Yet, incentives could still *indirectly* improve high-creativity production by priming the creative process, with gains to follow later.

[1] We secured approval from the institutional review boards at the two universities at which we conducted our experiments.

Our study adapts the "rebus puzzle" task that has been used in prior experimental studies of how incentives affect creativity (Kachelmeier, Reichert, and Williamson 2008; Kachelmeier and Williamson 2010; Erat and Gneezy 2016). Experimental instructions 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 several illustrative examples. The task requires participants to design rebus puzzles, which are later evaluated for creativity by independent raters who are blind to the treatment conditions. The common theme from prior studies using this task is that incentives can prompt greater effort and more ideas overall without lowering the number of high-creativity ideas, but there is no evidence from these studies that incentives can actually *improve* creativity.

The key difference characterizing our study is that we add a second stage that is separated in time from the initial experimental task, thus providing the detached incubation that realizes the gains from initial creative preparation (Dodds, Ward, and Smith 2012; Gilhooly 2016). In our first experiment, we implement four incentive compensation conditions initially: (1) quantity incentives, irrespective of creativity, (2) incentives for high-creativity ideas, (3) incentives for ideas that achieve a minimally acceptable creativity rating, and (4) a control condition with fixed pay. Relative to fixed pay, the only creativity difference we find in the first-stage task is that participants with a minimally acceptable creativity threshold generate significantly *fewer* high-creativity ideas. Although quantity incentives lead to more first-stage puzzles in total, no incentive condition outperforms fixed pay with respect to first-stage high-creativity puzzles. Ten days later, when participants return to collect their compensation, we introduce a second-stage task by asking participants to generate any additional rebus-puzzle ideas that come to mind. We do not implement any performance-contingent incentives in this second stage, instead providing all participants with a fixed stipend of $10 as a goodwill gesture for completing the second-stage request. Yet, we observe a treatment effect from *first-stage* incentives on *second-stage* high-creativity production. Specifically, participants with first-stage quantity incentives submit more high-creativity puzzles in the second stage than do participants with first-stage fixed pay. These findings are consistent with the reasoning that quantity incentives enrich the preparation stage of the creative process, thus leading to a creativity advantage after incubation.

Informed by our first experiment, we design a second experiment to generalize and extend these findings. Unlike Experiment 1, Experiment 2 operationalizes explicit long-term incentives, as participants learn in the first stage that the same compensation scheme will continue in the second stage. We consider only quantity-based pay and fixed pay in Experiment 2, as there is no evidence from Experiment 1 that incentives for high-creativity or a minimum creativity threshold do any better than simple quantity incentives in either the first or second stage. From a theoretical perspective, quantity incentives come closest to capturing the notion that incentives cannot force creativity instantly, but rather are best used to facilitate the process that benefits from creative incubation. Another key difference in Experiment 2 is that we split the two experimental stages by only 20 minutes. During this period, an experimenter escorts participants on a pleasant walk around a campus green space, with the intent of allowing participants to relax and take a break from the task. Despite the different setting and much shorter incubation period, our findings from Experiment 2 are similar to those from Experiment 1: participants with quantity-based incentives generate significantly more high-creativity puzzles than their fixed-pay counterparts only in the second stage of the experiment. Thus, even a 20-minute break appears sufficient to allow incentives to improve creativity, whereas the lack of any similar effect in the first stage of either Experiment 1 or Experiment 2 suggests that *some* break is necessary.

We corroborate these conclusions by conducting mediation analyses to identify the reasons why quantity incentives lead to an eventual creativity advantage. Consistent with the reasoning that incentives enhance creative preparation by encouraging experimentation with more flexible knowledge structures, our strongest candidate for mediation comes from a measure of *divergent thinking* that we proxy by counting the number of first-stage puzzles that depart from the patterns illustrated in instructional examples. We find that participants with quantity-based incentives produce significantly more divergent ideas in the first stage, likely because their desire to generate as many ideas as possible motivates them to seek new patterns. These divergent ideas are not necessarily creative in the sense of appealing to our creativity raters, but they do differ from the norm. To some extent in Experiment 1 and particularly in Experiment 2, we find evidence that first-stage divergent thinking mediates the second-stage creativity advantage of quantity incentives. We obtain stronger mediation evidence from divergent-thinking puzzles than from other puzzles, consistent with the reasoning that thinking differently is a prerequisite to thinking creatively. In contrast, we find no evidence of mediation from a self-reported measure of task enjoyment, suggesting that quantity-incentivized workers do not necessarily derive more enjoyment from a creative task, but they do appear to be better prepared.

Our cumulative evidence suggests that a likely contributor to the mixed findings on the effects of incentives on creativity in the psychology, management, economics, and accounting literatures is that experiments are short-term by nature. Hence, to the extent that incentives cannot force creativity immediately, they are unlikely to be effective in a one-shot experiment. We mitigate this problem by splitting our experiments into two stages, with an opportunity for creative incubation in between. Whether this incubation takes place over ten days in Experiment 1 or 20 minutes in Experiment 2, our findings indicate that performance-based incentives work best by priming the creative process, generating a greater number of initial divergent ideas that lead to an eventual advantage in creative ideas.

More generally, our study contributes to recent advances in the role of performance-based incentives in settings for which good performance requires effective strategies, not just raw effort (e.g., Farrell, Kadous, and Towry 2012; Choi, Hecht, and Tayler 2012, 2013). Incentives in such settings must do more than just reward 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 can prime the preparation-incubation-illumination process that culminates in more creativity.

Section II develops our theory in greater detail. Section III explains our design and results for Experiment 1. Section IV does the same for Experiment 2. Section V concludes.

II. LITERATURE REVIEW AND HYPOTHESIS 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 business (e.g., Fairbank and Williams 2001; Fallon and Senn 2006). In accounting, Grabner (2014) reports field evidence that creativity-dependent firms are more likely to utilize performance-contingent incentives, although her study does not identify the specific incentives these firms adopt or how incentives affect creativity. Thus, although Grabner's (2014) findings suggest that creativity can be *consistent* with performance-based incentives, the question remains open as to whether and how incentives can *improve* creativity.

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

(Byron and Khazanchi 2012).[2] Psychologists often advance the argument that creativity cannot be forced by incentives, and may in fact be undermined by extrinsic incentives that crowd out any intrinsic motivation to be creative (Amabile 1996). The literature supporting this view, however, generally considers creativity as an isolated task rather than the broader notion of creative *production*. We focus the current study on high-creativity production in recognition that businesses need a sustained development of new creative ideas to succeed. Accordingly, we build on prior experimental research by Kachelmeier et al. (2008) (hereafter KRW), in which participants design "rebus puzzles." This task confers the advantage of meaningful variation in both the quantity and creativity of ideas generated. Relative to creativity-based compensation or fixed pay, KRW find that simple quantity-based incentives significantly increase the total number of ideas generated without lowering the number of high-creativity ideas.

Although KRW's findings suggest that quantity-based compensation does not *harm* high-creativity production, their study does not find that quantity-based compensation *improves* high-creativity production either. In a follow-up experiment, Kachelmeier and Williamson (2010) find some self-selection benefits when participants can choose their desired pay scheme, but those who choose quantity-based pay end up generating no more (and no fewer) total high-creativity ideas than do other participants. Most recently, an experiment in the economics literature by Erat and Gneezy (2016) extends KRW's rebus-puzzle design task to a variety of creativity-based incentive schemes.[3] The authors reach the similar conclusion that incentives do not improve creativity, which they attribute to a "choking hypothesis."

[2] While research on creativity incentives is predominantly experimental, Byron and Khazanchi (2012) also consider several field and survey-based studies (see their Table 2), again with mixed findings. Also see a recent forum in the *Journal of Management Accounting Research* on field studies of controls in creative settings (Cools, Stouthuysen, and Van den Abbeele 2017; Chen 2017; Davila and Ditillo 2017; Speckbacher 2017).

[3] In a subsequently published erratum, Erat and Gneezy (2017) acknowledge that their experiment draws on task instructions developed by KRW.

Creativity as a Process

We posit that a prime reason for the mixed incentive effects reported in experiments to date is that creativity is not instantaneous, but rather involves a multistage process. The process model of creativity is often attributed as far back as a seminal book by London School of Economics cofounder Graham Wallas (1926). Over the ensuing decades, Wallas' stage-based structure has stood the test of time as a model of choice in both academic (e.g., Armbruster 1989; Csikszentmihalyi and Sawyer 2014; Gilhooly 2016) and practice-oriented (e.g., Stillman 2014) discussions of creativity. Our study's primary contribution is to examine how incentives influence creativity within the context of the creative process.

The first stage of the creative process is *preparation*, which Csikszentmihalyi (1996, 79) describes as "becoming immersed, consciously or not, in a set of problematic issues that are interesting and arouse curiosity." In simple terms, creative preparation gathers the raw materials of creativity, while not achieving creativity. Weisberg (1993, 45) goes so far as to describe the preparation stage of the creative process as a "period of intense conscious work, without success." The reason preparation is insufficient for creativity is that creative insight cannot be forced by effort alone (Amabile 1996), but rather emerges only after reflection and, as we discuss next, incubation. Still, experts in the area tend to agree that preparation, while insufficient for creativity, remains a necessary prerequisite to an effective creative process. In their review of studies of creative incubation, Dodds, Ward, and Smith (2012, 261) observe that "one advantage of [creative] preparation is to break up mental sets." Along the same line, Armbruster (1989, 178) asserts that the preparation stage of creativity is most effective when it builds a "flexible knowledge representation" in which "fragments of knowledge … can be moved about and reassembled into new knowledge structures."

These arguments relate to the literature on *divergent thinking* (e.g., Baer 1993; Runco and Acar 2012), in which people generate unconventional approaches to a problem. Runco and Acar (2012, 73) clarify that "divergent thinking is not synonymous with creativity," but it establishes a "meaningful potential." In simple terms, divergent ideas differ from the norm, but creativity requires ideas that are both different *and effective* (Runco and Jaeger 2012). A plausible role for incentives is to motivate people to think differently. Even simple quantity incentives can achieve this objective because the goal of maximizing ideas can incentivize people to try new approaches. Yet, as we discuss next, divergent thinking alone is unlikely to result in creative insight without the next stage of the creative process, incubation.

Weisberg (1993, 45) describes the incubation stage of the creative process as a period during which the creative task is "put aside and not thought about consciously." Theories of *why* incubation is helpful to creativity remain open to debate,[4] but the cumulative evidence indicates that detached "incubation periods, whether delayed or immediate, do have beneficial effects" (Gilhooly 2016, 3). Prior studies have implemented incubation periods in various ways, but a common theme is that effective incubation necessitates time away from the task rather than just additional time devoted to the task (Dodds et al. 2012).[5] Although we capture this feature over ten days in Experiment 1 and just 20 minutes in Experiment 2, both experiments separate the incubation period from the creative task. Our study is not designed to assess the benefits of

[4] As explained by Gilhooly (2016), alternative theories of the value of incubation involve (1) intermittent conscious thought that leverages the effectiveness of post-incubation creative activity, (2) "beneficial forgetting" that allows people to escape conventional mental representations, and (3) unconscious mental work that involves "active, but unconscious, or intuitive processing" (2016, 3). Gilhooly's (2016) review favors the third explanation over the first two, but this distinction is beyond the scope of our study, the more basic objective of which is to examine the effect of pre-incubation incentives on post-incubation creativity.

[5] In contrast, Erat and Gneezy (2016) find that it does not matter whether they give participants ten minutes or one hour to complete their task. Because they do not separate the additional time from the creative task, they do not incorporate an incubation stage as that term is used in the creative process literature.

incubation *per se*, but rather is designed to assess whether incentives in the preparation stage of the creative process can enhance the effectiveness of creative incubation.

The incubation stage is followed by *illumination* (e.g., Armbruster 1989) or *insight* (e.g., Csikszentmihalyi and Sawyer 2014), which realizes the creative fruits of the preparation and incubation stages.[6] We capture this notion by obtaining our primary dependent variable from a second-stage creative task, after incubation. To the extent that incentives facilitate the creative *process* rather than prompting creativity instantaneously, we should observe incentive effects only in the second-stage task, not at the time the first-stage incentives are implemented.

Incentivizing Creative Preparation

If the creative process starts with exploratory preparation, a role for incentives emerges. We first consider piece-rate quantity incentives as the simplest way to motivate as many initial ideas as possible. As KRW find from their one-shot experiment, participants with quantity incentives generate significantly more ideas than do those with fixed pay, although this advantage does not extend to the number of *high-creativity* ideas. From a creative process perspective, however, the goal is not to achieve creativity initially, but rather is to develop the initial preparation that can benefit from creative incubation. Similar to the claimed benefits of brainstorming (e.g., Paulus and Yang 2000), the additional ideas stimulated by quantity-based pay could serve this purpose. *New Yorker* cartoonist Matt Diffee explains this rationale in commenting on KRW as part of a "TedX" video, observing that the strategy he follows to generate creative cartoons emphasizes "quantity over quality" (Diffee 2013).[7]

[6] The creative process model also includes a final stage, *verification*, in which creative ideas from the illumination stage are refined and implemented (e.g., see Armbruster 1989, 180-181). We do not consider this stage in our study, but the role of incentives in creative verification could be a fruitful avenue for future research.
[7] 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.

In our first experiment, we also consider other possible incentives as alternatives to quantity-based pay. If the ultimate goal is high-creativity production, an obvious alternative is to pay participants only for high-creativity ideas. As noted previously, however, creativity does not necessarily emerge from raw effort (Amabile 1996), which is related to Erat and Gneezy's (2016) "choking" hypothesis. KRW observe far fewer submissions overall and no more high-creativity submissions in their condition that rewards participants for average creativity ratings. We modify this condition by rewarding the *number* of high-creativity puzzles rather than average creativity. Whether any gains from a high-creativity threshold outweigh any loss in the overall number of ideas generated is an empirical question that our first experiment addresses.

Finally, we consider an incentive scheme that imposes a minimum creativity rating to be eligible for payment. The rationale is to lessen the pressure of a scheme that rewards only high-creativity ideas, while at the same time deterring an abundance of low-creativity ideas that could arise from quantity incentives. Although this compromise is intuitively appealing, it nevertheless imposes a restriction that could impede the flow of ideas. As with our high-creativity incentive condition, our experimental design allows us to assess this possibility as an empirical question.

Our study is motivated by the potential for incentives in our first-stage task to lead to more beneficial incubation and hence greater high-creativity production in our second-stage task. Notwithstanding this reasoning, we acknowledge that the literature on how incentives affect creativity is unsettled (Byron and Khazanchi 2012), compelling us to state our primary hypothesis in the null form and to report two-tailed statistics:

Hypothesis (null form): Participants with first-stage performance-contingent compensation based on (1) the total quantity of ideas, (2) high-creativity ideas, or (3) ideas that meet a minimum creativity threshold will not exhibit different *second-stage* high-creativity production than participants with first-stage fixed pay.

III. EXPERIMENT 1

First-Stage Task and Design

A total of 104 volunteers recruited from undergraduate business classes participate in Experiment 1.[8] Their first-stage experimental task is to design rebus puzzles for 20 minutes. The experimental instructions 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 eleven illustrative examples obtained from public sources (e.g., Morris 1983; Griggs 2000). Except as noted below, our task instructions follow those reproduced in the Appendix to KRW (2008, 368-372).

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 "output box" to facilitate this process. Across conditions, participants are informed that a panel of creativity raters will evaluate their submitted puzzles afterwards for creativity on a ten-point scale, "where creativity refers to puzzles that are original, innovative, and clever" (quoted from the instructions). While KRW informed their participants that the researchers value quantity and creativity, our instructions focus specifically on *high-creativity production*, informing participants across conditions that "we value the *number of high-creativity 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 high-creativity puzzles as possible. Figure 1 illustrates a sampling of rebus puzzles submitted by our participants.

[8] 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.

< INSERT FIGURE 1 >

We randomly assign four versions of our instructions to operationalize the different incentives we test. The instructions note that different participants have different instructions, but do not inform participants of the nature of these differences. Participants in the *quantity* condition are informed that we will "*count the number of rebus puzzles you submit, no matter what creativity ratings those puzzles receive*" (emphasis in original). The instructions further inform participants that we will determine a cash payment rate to result in $45 for the participant in this condition submitting the most rebus puzzles 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 enables us to hold constant the *amount* of average compensation across conditions, while varying the *nature* of that compensation.

Participants in the *high-creativity* incentive condition are informed that they will be compensated for puzzles that receive a composite creativity rating of six or higher by the rating panel. In contrast to the "average creativity" and "creativity-weighted" conditions in KRW, 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 condition, participants in the high-creativity condition are informed that the participant submitting the most (least) puzzles rated six or higher will receive $45 ($5), thus generating average compensation of approximately $25.

Our third treatment, operationalizing a *minimum-creativity threshold*, is identical to the structure of the high-creativity condition except that it rewards puzzles rated *four* or higher.[9] The intent of this condition is to avoid the pressure of a high-creativity threshold while at the same time deterring ideas that clearly lack creativity. As with our other performance-based

[9] Similar to the instructional statement that approximately 15 percent of puzzles in prior experiments received a creativity rating at or above 6, the instructions also inform all participants that approximately 70 percent of puzzles in prior experiments received a creativity rating at or above 4.

compensation treatments, 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 approximately $25.

Our fourth condition is an experimental control with fixed pay of $25. Thus, we hold constant the *presence* of compensation at the same average amount ($25) as well as the same instruction that the experimenters "value the number of high-creativity puzzles you can construct." However, fixed-pay participants have no performance-based incentive to maximize either quantity or creativity.

We conduct the first-stage task in a controlled setting on the same day. Across conditions, the instructions note that participants will need to return ten days later to collect payment, allowing us to determine separate compensation pools for each condition under that condition's rules. We ask participants to wait ten days for payment even in the fixed-pay control condition, which the instructions explain by noting that "different versions of the research require waiting, and we want to pay all participants at the same time." To confirm understanding, a pre-experimental question asks participants to explain briefly in writing how they will be compensated, which an experimenter checks by visiting each participant's station to verify accurate responses before starting the task period.

Second-Stage Task

On the specified date, we provide a window of several hours for participants to return to collect their first-stage compensation. Upon arrival, an experimenter provides each participant with a packet containing an additional 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.

The instructions proceed with a request for at least one and up to ten additional creative rebus puzzles,[10] at the participants' discretion, along with a post-experimental questionnaire. The packet includes eleven blank index cards for participants to use for their additional rebus puzzles, which participants place in a large envelope along with any unused index cards. Although participants could have collected their first-stage compensation without doing the second-stage task, all participants agreed to our offer of $10 for completing this additional request. After handing in their envelopes, participants receive their first-stage compensation in the manner communicated in their respective conditions plus $10 for the second-stage task, as promised.

Two points are important to note about the second-stage procedures. First, to minimize the influence of first-stage outcomes on second-stage performance, participants complete the second-stage task before learning the amount of their first-stage compensation, with the exception that fixed-pay participants know they will be paid $25. Second, participants have no prior knowledge of the second-stage request for additional rebus puzzles until that request is made. This design choice is necessary because if participants had known in the first stage that we would ask for additional creative puzzles ten days later, they might have been tempted to "cheat" in the ensuing days by searching for rebus puzzles that would not have been their own creations. We pay Experiment 1 participants an additional fixed stipend of $10, held constant across conditions, as a goodwill gesture to compensate them for an additional request of which they were previously unaware. Thus, Experiment 1 does not manipulate long-term incentives, a point

[10] More specifically, the instructions direct participants to submit one "most creative" puzzle, along with up to ten additional puzzles, at the participant's discretion. Our primary dependent variable is high-creativity production, defined as the total number of puzzles submitted with a composite rating of six or higher. Tests of the creativity rating attained by each participant's most creative puzzle do not yield any treatment effects beyond those reported.

we address in Experiment 2 in which we maintain experimental control throughout by implementing a much shorter incubation period.

Creativity Ratings

To measure creativity, we elicit creativity ratings from independent panels of eight raters each. Creativity raters are student volunteers from business honors classes that do 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 the creativity of each puzzle, projected one at a time, on a 1 (lowest) to 10 (highest) scale. To minimize fatigue, we limit each panel to evaluating approximately 500 rebus puzzles. We randomize the order of puzzles for rating purposes, and raters are blind to treatment conditions. To ensure similar initial calibration across panels, each rating session begins with the same set of 40 puzzles.

We take several steps to reduce noise in our dependent variable, which is inherently subjective. First, within each panel, we correlate each rater's scores with panel averages and eliminate the rater with the lowest correlation. Second, we apply a calibration adjustment to each remaining rater by adding or subtracting to each rating the difference between that rater's overall average and the panel's overall average. This step ensures that the threshold for "high-creativity" puzzles is consistent across raters. Third, we drop the highest and lowest individual ratings for each puzzle to reduce the effect of outliers. We then calculate each puzzle's average rating by the five remaining panelists as our measure of creativity.

Experiment 1 Results – First Stage

Although our hypothesis is directed to second-stage high-creativity production, we begin our analysis with first-stage results for completeness. Table 1 shows means and standard deviations by condition. We use Analysis of Covariance (ANCOVA) for all statistical analyses, with covariates for age and gender.[11] Beginning with simple quantity counts, the ANCOVA in Table 2, Panel A shows a significant omnibus treatment effect for the number of first-stage puzzles submitted ($F_{3,98} = 12.50$; $p < 0.01$). In turn, the simple effect tests in Table 2, Panel B indicate that the quantity incentive condition (24.2 average total output) differs from the fixed-pay control condition (13.5 average; $F_{1,98} = 24.13$; $p < 0.01$), whereas the high-creativity and minimum-creativity incentive conditions (averages of 14.3 and 12.9, respectively) do not differ significantly from the 13.5 average in the fixed-pay condition ($p = 0.78$ and 0.70, respectively).[12] Thus, similar to KRW, we find that quantity incentives increase the total number of ideas participants generate in a creative design task.

< INSERT TABLES 1 AND 2 >

Turning to first-stage *high-creativity* production, defined from the instructions as the number of submitted puzzles with a composite creativity rating of six or higher,[13] we reach a different conclusion. Although the omnibus treatment effect in Table 2, Panel C is marginally

[11] Although random assignment to treatment conditions guards against any confounding influence of age or gender, these effects add variation to our dependent variable that ANCOVA serves to control. For gender, our data indicate that women significantly outperform men in Experiment 1 for second-stage quantity ($p = 0.016$), and in Experiment 2 for first-stage quantity ($p = 0.074$), second-stage quantity ($p = 0.117$), and second-stage high-creativity production ($p = 0.032$). For age, older students exhibit lower second-stage high-creativity production in Experiment 2 ($p = 0.065$). Age and gender effects are otherwise not significant at conventional levels, but we include both covariates in all reported analyses for consistency. Analyses without these covariates reach the same statistical conclusions, with the exception of a more marginal significance level for second-stage high-creativity production in Experiment 2, as we document later.

[12] All reported statistics are two-tailed.

[13] A high-creativity threshold of six is consistent with our *ex ante* representation to the participants that any puzzle rated six or higher would be considered a high-creativity puzzle, based on prior research (i.e., KRW) indicating that approximately 15 percent of prior participants' rebus puzzles attained such a rating. Reasonably consistent with this representation, our data indicate that 14.1 (16.5) percent of the first-stage puzzle submissions in Experiment 1 (2) achieve an average creativity rating of six or higher.

significant ($F_{3,98} = 2.28$; $p = 0.08$), the simple-effect tests in Table 2, Panel D indicate that the only significant pairwise difference is from comparing the minimum-creativity and fixed-pay conditions, for which the minimum-creativity threshold leads to significantly *fewer* first-stage high-creativity puzzles (averages of 1.40 and 2.69, respectively; $F_{1,98} = 4.70$; $p = 0.03$). A plausible interpretation is that a minimum-creativity threshold calls attention away from the desirability of high-creativity ideas. The broader point, however, is that no incentive condition, including quantity incentives, significantly outperforms fixed pay in terms of participants' high-creativity production at the time the incentives are initially applied.

Experiment 1 Results – Second Stage

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 structure of the two sessions. That is, the first stage is conducted as a formal experiment with 20 minutes dedicated to the task. The time participants devote to our second-stage request, in contrast, is purely at their discretion, with no advance notice. Participants could complete the second-stage request quickly 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.

Within conditions, total second-stage quantity averages are 2.93, 2.77, 1.48, and 2.00 in the quantity, high-creativity, minimum-creativity, and fixed-pay conditions, respectively (see Table 1). The ANCOVA for second-stage quantity in Table 3, Panel A does not reach statistical significance at conventional levels for an overall treatment effect ($F_{3,98} = 1.79$; $p = 0.15$), suggesting that any effect of first-stage incentives on second-stage total quantity is modest. Consistent with this conclusion, the simple effects reported in Table 3, Panel B indicate that none of the incentive conditions significantly differ from the fixed-pay control condition for the total number of second-stage puzzles submitted.

< INSERT TABLE 3 >

Notwithstanding the lack of a treatment effect on second-stage total quantity, our hypothesis is directed to second-stage *high-creativity* production. As the descriptive statistics in Table 1 indicate, participants with quantity incentives in the first stage submit an average of 1.48 second-stage puzzles that meet the high-creativity threshold, which exceeds the averages in the high-creativity (1.04), minimum-creativity (0.40) and fixed-pay (0.69) conditions. The ANCOVA in Table 3, Panel C indicates a significant overall treatment effect ($F_{3,98} = 2.62$; $p = 0.055$). In the simple-effect comparisons reported in Panel D, only the quantity condition is significantly greater than the fixed-pay control condition ($F_{1,98} = 4.46$; $p = 0.037$).[14] Thus, evidence indicates that initial quantity incentives confer creativity benefits ten days later. This result does not merely reflect the propensity of participants with quantity incentives to submit more puzzles, as these participants were no more creative than their fixed-pay counterparts when the incentives were in place. Rather, only in the second stage of the experiment do we find that quantity incentives can improve not just overall production, but also *high-creativity* production.

Thus, if asked to pick a winner among the incentives we compare to fixed pay in Experiment 1, we would identify the quantity condition. In untabulated analyses, the quantity condition significantly outperforms the second-stage high-creativity production of a composite blend of the three other conditions ($F_{1,98} = 5.78$; $p = 0.018$). In individual comparisons, the

[14] The second-stage quantity vs. fixed-pay comparison for high-creativity production remains statistically significant in Experiment 1 with a more stringent high-creativity rating threshold of seven ($p = 0.020$).

quantity condition does not significantly differ from the high-creativity condition ($F_{1,98} = 1.31$; $p = 0.25$), so we cannot necessarily claim that quantity incentives are more effective than high-creativity incentives for stimulating eventual high-creativity production. However, quantity incentives are clearly no *worse* than high-creativity incentives either (see Table 1), and as noted above, second-stage high-creativity production does not significantly differ between the high-creativity and fixed-pay conditions. When one also takes into account the relative simplicity of quantity incentives, the quantity condition emerges as the most promising candidate for an easily implementable way to stimulate the creative process, as we test further in Experiment 2.

Experiment 1 Results – Mediation Tests

In their attempt to define creativity from a synthesis of the literature, Runco and Jaeger (2012) explain that creative ideas must be both original (i.e., different) and effective (i.e., appealing). Runco and Acar (2012) develop these notions further in discussing divergent thinking as a precursor to creativity. In essence, a divergent idea is different, but is not necessarily effective or creative. To apply this reasoning to our study, we define a "divergent idea" as a submission that does not merely extend the pattern illustrated in one of the instructional examples. For instance, the submitted puzzle "too close for comfort" in Panel C of Figure 1 is a simple extension of the "too funny for words" example in the instructions, as both puzzles use the words "two" and "four" as homonyms to represent "too" and "for," writing out the first word two times and the second word four times. We establish specific criteria to identify extensions for all instructional examples.[15] Submissions *not* meeting these criteria are defined as

[15] 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);

divergent ideas, whereas we classify all other submissions as nondivergent ideas. For example, the submission "stop watch" in Panel C of Figure 1, in which the word "stop" appears in a watch dial, was not viewed by our raters as being particularly creative (composite rating of 4.04). Nevertheless, it qualifies as a divergent idea because it is unlike any of the instructional examples. From a theoretical perspective, divergent ideas serve as a logical candidate for mediating our findings because such ideas enhance creative preparation by establishing a flexible representation of the creative task (Armbruster 1989; Dodds et al. 2012).

Table 1 shows that participants with first-stage quantity incentives submit substantially more first-stage divergent ideas (average of 16.7) than do participants in any of the other conditions (averages of 10.6, 9.1, and 9.7 for the high-creativity, minimum-creativity, and fixed-pay conditions, respectively). To test the extent to which this first-stage difference in divergent ideas mediates the *second-stage* advantage of quantity-incentivized participants in high-creativity production, we employ the bootstrapping-based approach proposed by Preacher and Hayes (2004) and Hayes (2018, Ch. 3) that has gained considerable traction in the social sciences (e.g., see data from psychology reported by Hayes and Scharkow 2013, 1919).[16] We

(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).

[16] A simple mediation model is based on three relations: (1) $Y = i_1 + cX + e_1$; (2) $M = i_2 + aX + e_2$; and (3) $Y = i_3 + c'X + bM + e_3$, where Y is the dependent variable, X is the independent variable or factor, M is the proposed mediator, i_1, i_2, and i_3 are intercept terms, and e_1, e_2, and e_3 are error terms. The classic Baron and Kenny (1986) approach, which yields similar conclusions for our study, *infers* mediation from the statistical significance of c in (1), a in (2), and b in (3), along with the *lack* of significance of c' in (3). This approach has been criticized on the grounds that it is "not based on a quantification of the very thing it is attempting to test—the intervening effect" (Hayes 2009, 410). Critics point out that the product of the coefficients a from (2) and b from (3) is a more direct estimate of the extent to which X indirectly influences Y through the mediator M (e.g., see Zhao, Lynch, and Chen 2010). The Preacher and Hayes (2004) approach, as recently reiterated by Hayes (2018, Ch. 3), tests the significance of the product ab by resampling (with replacement) a large number (we use 1,000) of equally sized pseudo samples from the observed data, using the logic of bootstrapping to construct an empirical confidence interval from the resampled estimates of ab. Mediation is significant at the α level if the $1 - \alpha$ confidence interval does not include zero. Among others, Preacher and Hayes (2004) and Zhao et al. (2010) argue that the bootstrapping approach is superior to the parametric test of ab proposed by Sobel (1982) because the latter is hampered by the nonnormal distribution of Sobel's test statistic.

find from a 95 percent bootstrapped confidence interval that first-stage divergent ideas significantly mediate the overall difference in second-stage high-creativity production between the quantity condition and a composite of the other three conditions. However, if we restrict the analysis to the quantity and fixed-pay conditions only, the confidence interval falls to 70 percent, which would generally not be considered as significant evidence of mediation. We return to this point in Experiment 2, in which our exclusive focus on the quantity and fixed-pay conditions enhances our ability to corroborate the role of divergent thinking as the conceptual driver of how initial quantity incentives lead to eventual creativity. Nevertheless, we obtain enough support in Experiment 1 from comparing the quantity condition against all other conditions to suggest that divergent thinking likely plays a key role in transforming productivity to creativity.

Given the focus in psychology and management on the tradeoff between intrinsic and extrinsic motivation (e.g., Amabile 1996; Deci et al. 1999), we also consider the potentially mediating effect of a self-reported measure of task enjoyment. 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, high-creativity, minimum-creativity, and fixed-pay conditions, respectively, do not statistically differ ($F_{3,98} = 1.42$; $p = 0.24$, untabulated). Indeed, the quantity incentive condition does not even have the highest mean, such that task enjoyment clearly does not mediate the positive effect of first-stage quantity incentives on second-stage high-creativity production. We conclude that quantity incentives do not necessarily lead participants to enjoy the task any more than they would with fixed pay, but it does make them more productive, more divergent, and eventually more creative.

IV. EXPERIMENT 2

Task and Design

We design Experiment 2 in the spirit of a "differentiated replication" of Experiment 1 (Lindsay and Ehrenberg 1993; Salterio 2014), with modifications intended to corroborate, generalize, and extend our Experiment 1 conclusions. As noted by Lindsay and Ehrenberg (1993, 220), "… the more explicit, differentiated, and/or deliberate such variations in the conditions of observation are while still obtaining the same result…, the more telling and exciting the outcome." Our primary modification in Experiment 2 is that we test explicit long-term incentives, meaning that participants in the first stage of the experiment know that the same incentive structure will continue in a second stage. This feature is made possible by a much shorter incubation period (i.e., 20 minutes rather than ten days), during which an experimenter escorts participants on a relaxing walk while maintaining control over the setting.

Participants in Experiment 2 are 63 student volunteers across three experimental sessions, recruited from an undergraduate business program at a university different from the university used in Experiment 1.[17] Participants complete the same rebus-puzzle task using the same task instructions and examples as in Experiment 1, including the same statement that "we value the number of high-creativity puzzles you can construct," defining "high-creativity" as puzzles that achieve a creativity rating of six or higher. One difference, however, is that the first-stage work period is 12 minutes in Experiment 2, as opposed to 20 minutes in Experiment 1. The only other substantive task difference is the explicit indication before participants begin the first stage that

[17] Although both universities used in our study have similar profiles and missions, we acknowledge the potential for subject pool differences. This should not be a problem for our research objectives, however, as we analyze each experiment separately. Indeed, as with other variations in Experiment 2, we gain additional assurance from reaching similar findings in different settings.

the entire experiment consists of two work periods with a break in between. Specifically, all participants in Experiment 2 read the following additional instructional passage:

> You will have a total of 24 minutes to construct rebus puzzles, which is comprised of two work sessions of 12 minutes each. After the first 12-minute work session, everyone will take a 20-minute walk outside of this classroom. Following this walk, you will construct rebus puzzles for the second 12-minute work session. The walk is an important aspect of our study. To help us maintain control of this experiment, please refrain from using your cell phone or talking to your fellow participants during the walk.

Following this instruction, participants read about their compensation, for which different versions are randomly assigned to participants as in Experiment 1. To increase power, we simplify the design in Experiment 2 by including only quantity incentives and a fixed-pay control condition. In Experiment 1, quantity-based pay is the only condition to differ significantly from fixed pay in second-stage high-creativity production, making it the most logical candidate to test again in Experiment 2. An added advantage of focusing only on quantity-based pay and fixed pay in Experiment 2 is that neither condition requires us to secure creativity ratings before paying participants, thereby enabling us to complete the experiment in a continuous block of time (including the incubation period) instead of waiting several days.

Compensation instructions in each condition are patterned after those in Experiment 1, except that we raise the average pay from $25 to $30, with anchors of $10 and $50 for the lowest and highest performers in the quantity condition and a fixed stipend of $30 in the fixed-pay condition. Importantly, participants in both conditions understand that their compensation is for *both* work sessions. That is, participants in the quantity incentive condition understand that they will receive additional payment for each puzzle submitted in either the first or second work session, added together to determine the participant's total quantity. Similarly, participants in the fixed-pay condition understand that they will be paid $30 for completing both work sessions.

After reading the task instructions and completing pre-experimental questions to ensure comprehension, participants begin the first-stage task, conducted in a manner similar to Experiment 1. After this 12-minute task ends, one of the experimenters escorts all participants on the announced walk. The 20-minute walk is around a large green space on the university campus, at a relaxed pace. The presence of an experimenter during the walk maintains control to guard against extraneous communication or other activities, while the physical act of walking outside separates this incubation phase from the creative design task. While the participants are away, a second experimenter counts each participant's first-stage puzzles and places a rubber band around them to separate this output from the second stage.

After completing the walk, participants return to the task room. We then begin the second 12-minute work period with no further instruction. As we discuss later, conducting 12-minute work periods in both stages enables a direct comparison across stages that we are unable to undertake in Experiment 1 because the settings of each stage differ in that experiment. After finishing the second stage, participants complete a post-experimental questionnaire while the experimenters count submissions and process payments in the manner described in each condition. Participants leave after collecting their cash compensation.

For analysis purposes, we obtain creativity ratings in a manner similar to that described for Experiment 1, calculating averages from rating panels comprised of business-honors students who did not participate in the experiment. Following the same approach as in Experiment 1, we drop the rater in each panel who is least correlated with the panel's average, apply a calibration adjustment for each panelist by adding or subtracting a constant to each rating based on the difference between each rater's overall average and the panel's overall average, and drop the high and low rating for each puzzle evaluated.

Experiment 2 Results – First and Second Stages

As with Experiment 1, we begin our analysis of Experiment 2 with first-stage production, with descriptive statistics tallied in Table 4 and ANCOVA results in Table 5, before turning to the second stage to test our hypothesis. Consistent with Experiment 1, Panel A of Table 5 confirms that participants with quantity incentives in Experiment 2 submit significantly more first-stage puzzles than do their fixed-pay counterparts (15.2 vs. 8.3; $F_{1,59} = 22.10$; $p < 0.01$). However, also consistent with Experiment 1, Panel B of Table 5 indicates that the advantage of quantity-incentivized participants over fixed pay in total first-stage production *does not* extend to high-creativity production (1.84 vs. 1.94; $F_{1,59} = 0.06$; $p = 0.80$). Thus, we successfully replicate the findings of Experiment 1 and prior studies such as KRW that quantity incentives increase total quantity but do not increase high-creativity production at the time those incentives are initially applied.

< INSERT TABLES 4 AND 5 >

After participants return for the second stage, those with quantity incentives continue to enjoy a quantity advantage over the fixed-pay condition (20.5 vs. 9.1; $F_{1,59} = 33.61$; $p < 0.01$), as documented in Table 5, Panel C. This time, however, quantity-incentivized participants also submit significantly more puzzles that clear the high-creativity threshold (2.90 vs. 2.06; $F_{1,59} = 3.48$; $p = 0.067$), as documented in Panel D.[18] We acknowledge that a *p*-value of 0.067 is at a borderline level, but given our Experiment 1 findings, using a two-tailed test in Experiment 2 is perhaps overly conservative. That is, a one-tailed *p*-value of 0.034 could be justified from the

[18] Unlike Experiment 1 (see footnote 14), the second-stage high-creativity advantage of quantity-incentivized participants is no longer statistically significant in Experiment 2 if we use a more stringent high-creativity rating threshold of seven rather than six. A threshold of six is more defensible *ex ante*, as it is consistent with the instructional statement to all participants that we "value the number of high-creativity puzzles you can construct," noting explicitly that "a rating at or above 6 would be considered a high-creativity puzzle." A high-creativity threshold of six also meets the intuitive criterion of a "tight, but attainable" standard, accounting for approximately 15 percent of puzzles submitted. In contrast, puzzles rated seven or higher account for only approximately 5 percent of the distribution.

reasoning that it is unlikely, given our theory and evidence from Experiment 1, that quantity incentives would worsen rather than improve second-stage high-creativity production. On balance, Experiment 2 corroborates our conclusions from Experiment 1. That is, incentives do not appear to benefit creativity when initially applied, but after an explicit incubation period even as short as 20 minutes in Experiment 2, a creativity advantage from quantity incentives emerges.[19]

Experiment 2 Results – Mediation Tests

As in Experiment 1, we test first-stage divergent ideas, defined the same way, as our most likely candidate for mediating the second-stage incentive effect we observe in Experiment 2. Table 4 shows that participants with quantity incentives submit more first-stage divergent ideas (average of 9.3) than do their fixed-pay counterparts (average of 5.4), a difference that is statistically significant ($F_{1,59} = 17.92$; $p < 0.01$). Thus, although quantity-incentivized participants are no more creative than fixed-pay participants in the first stage, they are more predisposed to try ideas that depart from the patterns illustrated in the instructional examples. A bootstrapping-based mediation analysis (Preacher and Hayes 2004; Hayes 2018, Ch. 3) finds that first-stage divergent ideas mediate the second-stage high-creativity advantage of quantity-incentivized participants at a 95 percent confidence interval, which is stronger than the mediation evidence we obtain from comparing the quantity and fixed-pay conditions in Experiment 1. We also test for mediation from first-stage *nondivergent* ideas (i.e., submissions that extend the pattern illustrated by an instructional example), finding that these ideas *do not* mediate our findings. Indeed, we reject the null hypothesis that divergent and nondivergent ideas serve as

[19] We also observe unusually strong age ($p = 0.065$) and gender ($p = 0.032$) effects in second-stage high-creativity production in Experiment 2, with better performance by women and younger participants (see footnote 11). Reflecting the variance contributed by these effects, if we remove the covariates for age and gender, the *p*-value for quantity incentives in the test of second-stage high-creativity production falls to 0.110 (0.055 one-tailed), or to 0.097 (0.049 one-tailed) if we include only a covariate for gender.

equivalent mediators at the $p < 0.05$ level. Thus, evidence suggests that quantity-incentivized participants' creativity advantage in the second stage can be attributed to their greater willingness to engage in divergent thinking in the first stage, as opposed to a more basic sense of inertia from a large volume of ideas in general.

For completeness, we also elicit and test the effect of self-reported task enjoyment. As in Experiment 1, we do not find a treatment effect for task enjoyment, with means of 4.87 and 5.19 on a seven-point Likert scale for the quantity and fixed-pay conditions, respectively ($F_{1,59} = 0.51$; $p = 0.47$, untabulated). Hence, consistent with our conclusions for Experiment 1, the benefit of first-stage quantity incentives on second-stage high-creativity production appears to be independent of the extent to which participants self-report that they enjoyed the task.

Supplemental Analysis for Experiment 2 – Stage × Treatment Interaction

The different task structure in Experiment 2 affords the benefit of allowing us to directly compare the two experimental stages. That is, both Experiment 2 stages are of the same duration and operate under the same incentive structure, in contrast to Experiment 1 in which we manipulate incentives only in the timed first-stage task, followed by an untimed request in the second stage that most participants completed quickly. Accordingly, as a supplemental analysis, we conduct an ANCOVA for Experiment 2 with the experimental stage as a repeated-measures factor and the incentive treatment as a between-participants factor. As illustrated in Figure 2, results (untabulated) indicate a significant stage × treatment interaction ($F_{1,59} = 3.91$; $p = 0.053$ two-tailed, or 0.026 one-tailed conditional on the expectation of a larger treatment effect in the second stage). Thus, we find statistical evidence that the influence of quantity incentives *differs* between the two stages, with a positive effect on high-creativity production emerging only after 20 minutes of relaxed incubation. We interpret these results as lending further support to the reasoning that incentives do not improve creativity instantaneously, but rather facilitate the creative *process*.

< INSERT FIGURE 2 >

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 subtle effects of incentives on production strategies for tasks with multiple dimensions that respond to effort in different ways (e.g., KRW 2008; Farrell et al. 2012; Choi et al. 2012, 2013). In two related experiments, we continue this progress by examining the effects of initial performance-based incentives on the process that generates creative ideas, using the "rebus puzzle" design task originated by KRW. The distinguishing characteristic of our study is that we examine the effects of incentives on high-creativity production at two points in time, separated by an explicit incubation period that the psychology literature indicates is an integral part of the creative process (e.g., Dodds et al. 2012; Csikszentmihalyi and Sawyer 2014; Gilhooly 2016). In Experiment 1, the incubation period occurs over the ten days between when participants complete the first-stage task and when they return to collect their payment. In Experiment 2, the incubation period is just 20 minutes, during which an experimenter escorts participants for a relaxing walk in between two compensated work periods. In both experiments, relative to a fixed-pay control condition, we find that a simple piece-rate compensation scheme for the quantity of ideas submitted yields more high-creativity ideas only in the second-stage task that occurs after the incubation period. Accordingly, our evidence suggests that incentives do not necessarily have a direct effect on creativity when they are implemented, but rather have an indirect effect by stimulating the creative process that benefits from incubation.

We corroborate this interpretation in mediation analyses that point to divergent thinking as the most likely conceptual driver of our findings. We proxy for divergent ideas by tallying the number of first-stage ideas that do not merely extend the patterns illustrated by examples of rebus puzzles in the instructional materials. In both experiments, we find that participants with quantity incentives submit significantly more divergent ideas in the first-stage task. We find partial evidence in Experiment 1 and stronger evidence in Experiment 2 that these initial divergent ideas mediate the advantage quantity-incentivized participants enjoy in second-stage high-creativity production, consistent with Runco and Acar's (2012) explanation that divergent thinking, while not synonymous with creativity, is a precursor to creativity.

In Experiment 1, we also test creativity-based performance incentives in two additional conditions that pay participants only for their high-creativity ideas or only for ideas that meet a minimum-creativity threshold. Neither of these conditions significantly outperforms fixed pay in second-stage high-creativity production. Accordingly, we focus only on the quantity-incentive and fixed-pay conditions in Experiment 2. From a creative-process perspective, quantity incentives confer the advantage of stimulating the free flow of initial ideas without restrictive blockage, similar to the benefits of creative brainstorming claimed by Paulus and Yang (2000) and illustrated in a practical context by Diffee (2013). That being said, we acknowledge that our study is not designed to fully differentiate the long-term advantages and disadvantages of quantity-based incentives versus creativity-based incentives, which is a question we leave to future research.

We further acknowledge that our study focuses on high-creativity *production* rather than creativity *per se*. That is, we count the number of submissions that meet a pre-specified high-creativity threshold to obtain our dependent variable, as opposed to measuring the level of creativity for one creative project (as in Chen, Williamson, and Zhou 2012, for example). We believe that the notion of creative production is central to businesses that need a sustained development of new creative ideas to succeed. Still, we acknowledge that the beneficial effect of quantity incentives that we detect for creative production may not generalize to creative tasks centered on a single project. We leave further consideration of this point to future research.

Our study helps to address conflicting views in the literature on the effectiveness of monetary incentives when tasks are defined by "soft" quality characteristics such as creativity. Although psychologists widely hold the view that explicit incentives crowd out 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 perspectives is that incentive effects are not instantaneous. Consistent with this theme, we encourage further exploration of incentive effects in multistage settings. As our Experiment 2 illustrates, even a 20-minute break can be enough to manifest incentive effects that might not be observable otherwise.

FIGURE 1

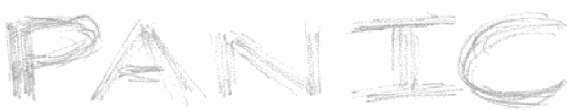
Examples of Rebus Puzzles Submitted by Participants[a]

Panel A: High-Creativity Puzzles

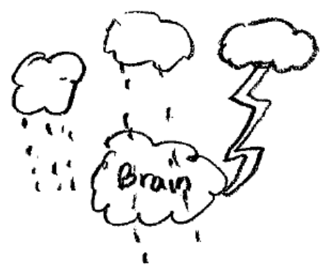
Solution: Keyboard
Creativity rating: 8.41

Solution: OK
Creativity rating: 8.96

Panel B: Borderline High-Creativity Puzzles

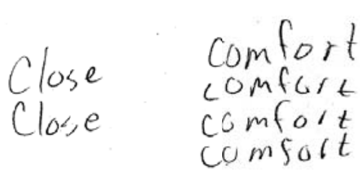


Solution: Widespread panic
Creativity rating: 6.21

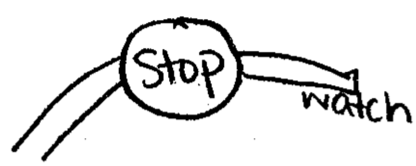


Solution: Brainstorm
Creativity rating: 6.10

Panel C: Moderate-Creativity Puzzles



Solution: Too close for comfort
Creativity rating: 4.47

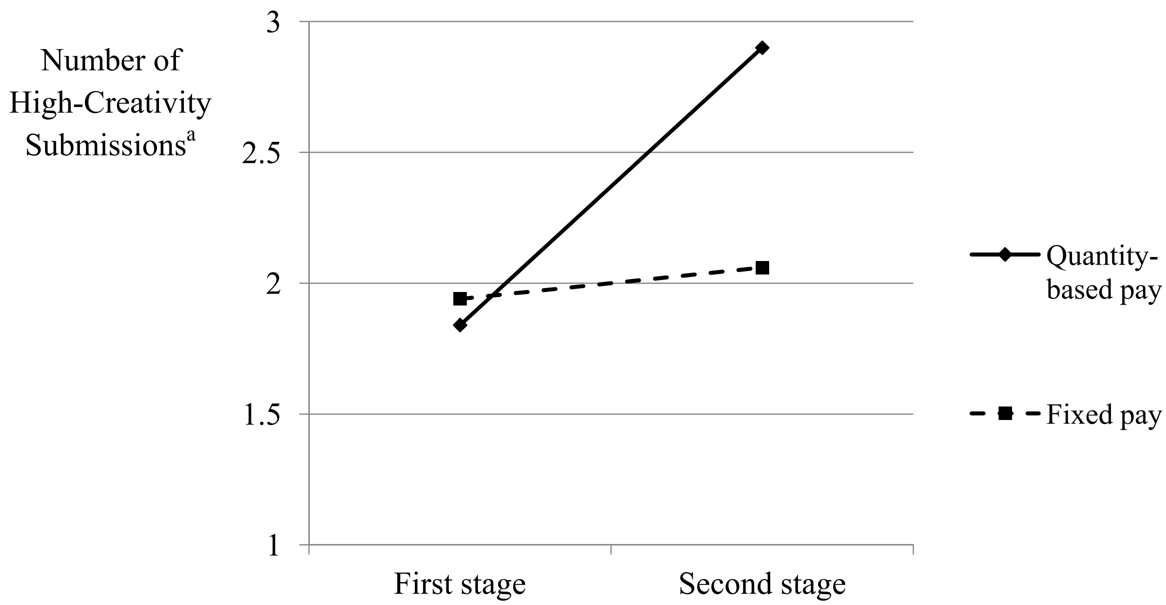


Solution: Stop watch
Creativity rating: 4.04

[a] Each panel illustrates a rebus puzzle submission from one of our Experiment 1 participants on the left side and from one of our Experiment 2 participants on the right side. The left and right examples are not intended to illustrate differences between the two experiments, but serve only as representative examples of output for our task.

FIGURE 2

Stage × Treatment Interaction for High-Creativity Production in Experiment 2



[a] High-creativity submissions are those that receive a creativity rating of six or higher on a ten-point scale, using the definition provided to participants in the experimental instructions.

TABLE 1

Means (Standard Deviations) for Experiment 1

	Condition				
	Quantity-Based Pay	High-Creativity Incentives	Minimum Creativity Threshold	Fixed Pay (Control)	Overall Averages
First-stage total quantity	24.15 (12.21)	14.27 (5.97)	12.92 (5.11)	13.50 (4.93)	16.21
First-stage high-creativity production[a]	2.26 (2.09)	2.77 (2.35)	1.40 (1.55)	2.69 (1.87)	2.28
First-stage divergent ideas[b]	16.67 (8.72)	10.62 (4.75)	9.12 (4.00)	9.69 (4.01)	11.53
Second-stage total quantity	2.93 (2.83)	2.77 (3.14)	1.48 (0.96)	2.00 (2.28)	2.30
Second-stage high-creativity production[a]	1.48 (1.93)	1.04 (1.75)	0.40 (0.58)	0.69 (0.88)	0.90
N[c]	27	26	25	26	

[a] High-creativity production reflects submissions that receive a creativity rating of six or higher on a ten-point scale, using the definition provided to participants in the experimental instructions.
[b] Divergent ideas 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 Experiment 1. Four participants, comprising one in the high-creativity incentives condition, two in the minimum-creativity threshold 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

Experiment 1: First-Stage ANCOVA Results[a]

Panel A: Omnibus ANCOVA for First-Stage Total Quantity

Source of variance:	df	MS	*F*	*p*-value[c]
Incentive conditions[b]	3	760.73	12.50	< 0.001
Error	98	60.88		

Panel B: Follow-up Comparisons to Fixed Pay for First-Stage Total Quantity

Source of variance:	df	MS	*F*	*p*-value[c]
Quantity-based pay vs. fixed pay	1	1468.78	24.13	< 0.001
High-creativity incentives vs. fixed pay	1	4.79	0.08	0.780
Minimum creativity threshold vs. fixed pay	1	9.08	0.15	0.700

Panel C: Omnibus ANCOVA for First-Stage High-Creativity Production[d]

Source of variance:	df	MS	*F*	*p*-value[c]
Incentive conditions[b]	3	9.24	2.28	0.084
Error	98	4.05		

Panel D: Follow-up Comparisons to Fixed Pay for First-Stage High-Creativity Production

Source of variance:	df	MS	*F*	*p*-value[c]
Quantity-based pay vs. fixed pay	1	2.33	0.58	0.450
High-creativity incentives vs. fixed pay	1	0.12	0.03	0.864
Minimum creativity threshold vs. fixed pay	1	19.04	4.70	0.033

[a] Although participants are randomly assigned to conditions, all reported analyses include age and gender covariates from the post-experimental questionnaire (untabulated) to control for variation in the dependent variable attributable to participant demographics. We reach the same statistical significance conclusions from this table without these covariates.

[b] The omnibus ANCOVA tests for overall differences across the four conditions we manipulate in the first stage of Experiment 1: quantity-based pay, high-creativity incentives, a minimum creativity threshold, and a fixed-pay control condition.

[c] All reported *p*-values are two-tailed.

[d] High-creativity production reflects submissions that receive a creativity rating of six or higher on a ten-point scale, using the definition provided to participants in the experimental instructions.

TABLE 3

Experiment 1: Second-Stage ANCOVA Results[a]

Panel A: Omnibus ANCOVA for Second-Stage Total Quantity

Source of variance:	df	MS	*F*	*p*-value[c]
Incentive conditions[b]	3	10.18	1.79	0.154
Error	98	5.68		

Panel B: Follow-up Comparisons to Fixed Pay for Second-Stage Total Quantity

Source of variance:	df	MS	*F*	*p*-value[c]
Quantity-based pay vs. fixed pay	1	13.79	2.43	0.123
High-creativity incentives vs. fixed pay	1	8.68	1.53	0.219
Minimum creativity threshold vs. fixed pay	1	0.93	0.16	0.687

Panel C: Omnibus ANCOVA for Second-Stage High-Creativity Production[d]

Source of variance:	df	MS	*F*	*p*-value[c]
Incentive conditions[b]	3	5.27	2.62	0.055
Error	98	2.01		

Panel D: Follow-up Comparisons to Fixed Pay for Second-Stage High-Creativity Production

Source of variance:	df	MS	*F*	*p*-value[c]
Quantity-based pay vs. fixed pay	1	8.97	4.46	0.037
High-creativity incentives vs. fixed pay	1	1.84	0.91	0.342
Minimum creativity threshold vs. fixed pay	1	0.45	0.23	0.636

[a] Although participants are randomly assigned to conditions, all reported analyses include age and gender covariates from the post-experimental questionnaire (untabulated) to control for variation in the dependent variable attributable to participant demographics. We reach the same statistical significance conclusions from this table without these covariates.
[b] The omnibus ANCOVA tests for overall differences in the second stage across the four conditions we manipulate in the first stage of Experiment 1: quantity-based pay, high-creativity incentives, a minimum creativity threshold, and a fixed-pay control condition.
[c] All reported *p*-values are two-tailed.
[d] High-creativity production reflects submissions that receive a creativity rating of six or higher on a ten-point scale, using the definition provided to participants in the experimental instructions.

TABLE 4

Means (Standard Deviations) for Experiment 2

	Condition		
	Quantity-Based Pay	**Fixed Pay (Control)**	**Overall Averages**
First-stage total quantity	15.23 (7.74)	8.31 (3.22)	11.77
First-stage high-creativity production[a]	1.84 (1.42)	1.94 (1.64)	1.89
First-stage divergent ideas[b]	9.32 (4.12)	5.38 (3.22)	7.35
Second-stage total quantity	20.52 (10.57)	9.06 (3.83)	14.79
Second-stage high-creativity production[a]	2.90 (2.15)	2.06 (1.97)	2.48
N	31	32	

[a] High-creativity production reflects submissions that receive a creativity rating of six or higher on a ten-point scale, using the definition provided to participants in the experimental instructions.
[b] Divergent ideas are submissions that do not exhibit one of the patterns illustrated in the instructional examples.

TABLE 5

Experiment 2 ANCOVA Results[a]

Panel A: ANCOVA for First-Stage Total Quantity

Source of variance:	**df**	**MS**	***F***	***p*-value[b]**
Quantity-based pay vs. fixed pay	1	751.36	22.10	< 0.001
Error	59	34.00		

Panel B: ANCOVA for First-Stage High-Creativity Production[c]

Source of variance:	**df**	**MS**	***F***	***p*-value[b]**
Quantity-based pay vs. fixed pay	1	0.15	0.06	0.803
Error	59	2.38		

Panel C: ANCOVA for Second-Stage Total Quantity

Source of variance:	**df**	**MS**	***F***	***p*-value[b]**
Quantity-based pay vs. fixed pay	1	2077.78	33.61	< 0.001
Error	59	61.82		

Panel D: ANCOVA for Second-Stage High-Creativity Production[c]

Source of variance:	**df**	**MS**	***F***	***p*-value[b]**
Quantity-based pay vs. fixed pay	1	13.51	3.48	0.067
Error	59	3.88		

---

[a] Although participants are randomly assigned to conditions, all reported analyses include age and gender covariates from the post-experimental questionnaire (untabulated) to control for variation in the dependent variable attributable to participant demographics. We reach the same statistical significance conclusions in Panels A, B, and C without these covariates. In Panel D, the *p*-value becomes 0.110 (0.055 one-tailed) if we remove the covariates for age and gender, or 0.097 (0.049 one-tailed) if we include only a covariate for gender.
[b] All reported *p*-values are two-tailed. The p-value for Panel D becomes 0.034 if we convert it to a one-tailed statistic conditional on our directional finding from Experiment 1.
[c] High-creativity production reflects submissions that receive a creativity rating of six or higher on a ten-point scale, using the definition provided to participants in the experimental instructions.

References

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

Armbruster, B. B. 1989. Metacognition in creativity. In *Handbook of Creativity*, edited by J. A. Glover, R. R. Ronning, and C. R. Reynolds, 177-182. New York: Plenum 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.

Baer, J. 1993. *Creativity and Divergent Thinking: A Task-Specific Approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Chen, C. X. 2017. Discussion of 'Management control for stimulating different types of creativity: The role of budgets.' *Journal of Management Accounting Research* 29 (3): 23-26.

Chen, C. X., M. G. Williamson, and F. H. Zhou. 2012. Reward system design and group creativity: An experimental investigation. *The Accounting Review* 87 (6): 1885-1911.

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.

Cools, M., K. Stouthuysen, and A. Van den Abbeele. 2017. Management control for stimulating different types of creativity: The role of budgets. *Journal of Management Accounting Research* 29 (3): 1-21.

Csikszentmihalyi, M. *Creativity: Flow and the Psychology of Discovery and Invention*. New York: HarperCollins.

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.

Davila, A., and A. Ditillo. 2017. Management control systems for creative teams: Managing stylistic creativity in fashion companies. *Journal of Management Accounting Research* 29 (3): 27-47.

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.

Dodds, R. A., T. B. Ward, and S. M. Smith. 2012. Incubation in problem solving and creativity. In *The Creativity Research Handbook*, vol. 3, edited by M. A. Runco, 251-284. New York: Hampton Press.

Erat, S., and U. Gneezy. 2016. Incentives for creativity. *Experimental Economics* 19 (2): 269-280.

Erat, S., and U. Gneezy. 2017. Erratum to: Incentives for creativity. *Experimental Economics* 20 (1): 274-275.

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.

Gilhooly, K. J. 2016. Incubation and intuition in creative problem solving. *Frontiers in Psychology* 7 (Article 1076): 1-9.

Grabner, I. 2014. Incentive system design in creativity-dependent firms. *The Accounting Review* 89 (5): 1729-1750.

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

Hayes, A. F. 2009. Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs* 76 (4): 408-420.

Hayes, A. F. 2018. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*, 2nd ed. New York: The Guilford Press.

Hayes, A. F., and M. Scharkow. 2013. The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analyses: Does method really matter? *Psychological Science* 24 (10): 1918-1927.

Kachelmeier, S. J., B. E. Reichert, and M. G. Williamson. 2008. Measuring and motivating quantity, creativity, or both. *Journal of Accounting Research* 46 (2): 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.

Lindsey, R. M., and A. S. C. Eherenberg. 1993. The design of replicated studies. *The American Statistician* 47 (3): 217-222.

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 (4): 717-731.

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.

Salterio, S. E. 2014. We don't replicate accounting research – or do we? *Contemporary Accounting Research* 31 (4): 1134-1142.

Sobel, M. E. 1982. Asymptotic confidence intervals for indirect effects in structural equation models. In *Sociological Methodology*, edited by S. Leinhart, 290-312. San Francisco, CA: Jossey-Bass.

Speckbacher, G. 2017. Creativity research in management accounting: A commentary. *Journal of Management Accounting Research* 29 (3): 49-54.

Stillman, J. 2014. *The Four Stages of Creativity.* Inc.com online webpage, available at https://www.inc.com/jessica-stillman/the-4-stages-of-creativity.html.

Wallas, G. 1926. *The Art of Thought*. New York: Harcourt Brace.

Weisberg, R. W. 1993. *Creativity: Beyond the Myth of Genius.* New York: W. H. Freeman and Company.

Zhao, X., J. G. Lynch, Jr., and Q Chen. 2010. Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research* 37: 197-206.