The Role of Incentives in Sustaining High-Creativity Production

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
michael.williamson@mccombs.utexas.edu
The University of Texas at Austin

March 2015

We appreciate comments on earlier versions from E.B. Altiero, Spencer Anderson, Tim Bauer, Jason Brown, Will Ciconte, Angelo Ditillo, Stephanie Grant, Andy Kitto, Lisa LaViers, Justin Leiby, Michael Majerczyk, Bill Messier, Sharon Noppe, Kathy Rupar, Tom Vance, Brian White, participants at the 2015 Management Accounting Section Midyear Meeting, and workshop participants at Bocconi University, the University of Florida, Georgia State University, the University of Graz, the University of Illinois at Urbana-Champaign, Indiana University, Maastricht University, the University of Nevada at Las Vegas, Technical University of Munich, the University of Texas at Arlington, and the University of Texas at Austin.

The Role of Incentives in Sustaining High-Creativity Production

Abstract: We examine the effects of performance-based compensation incentives on high-creativity production at two points in time: (1) a first-stage experiment during which the incentive manipulation is implemented, and (2) a follow-up second-stage event ten days after the removal of the first-stage incentives. We find that experimental participants receiving quantity-based pay for their first-stage production generate more high-creativity ideas *in the second stage* than do participants receiving fixed pay, even though we observe no compensation-based difference in high-creativity production in the first-stage experiment itself. These findings support the premise that, although incentives may have no immediate beneficial effect on creativity, those operating under performance-based incentives nevertheless achieve more progress on the task, thus helping them to sustain long-term creativity. Our results provide a possible explanation for the popularity of performance-based incentives even in creativity-dependent firms.

The Role of Incentives in Sustaining High-Creativity Production

I. Introduction

How performance-contingent incentives affect creativity remains a controversial question in psychology, management, and accounting (Byron and Khazanchi 2012). While an experimental approach to this question affords control over extraneous environmental forces, a key limitation is that experiments are generally short-term by design, measuring the influence of incentives on creativity only at the time of the experiment. Accordingly, the typical laboratory experiment does not capture the longer-term sense of "defocused attention" that neurobehavioral research indicates is essential to sustaining creative performance over time (Dietrich 2004). A longer horizon is also central to practice, in which recent survey research finds that firms that depend more on creativity also depend more on performance-contingent incentive schemes (Grabner 2014). Accordingly, our research objective is to maintain the benefits of experimental control, while extending the question of how incentives influence creativity to a longer horizon.

We achieve this objective by designing a two-stage experiment, in which we test the effects of initial performance-contingent incentives on creative production in a second-stage task that we invite participants to complete ten days later. Importantly, we do not merely test the effects of long-term incentives, as our participants are unaware of the second stage of the experiment when they complete the first stage, and we do not manipulate incentives in the second stage. Conceptually, testing the effects of long-term incentives would be "turning two dials at once," involving a longer task duration and a fundamental change to the incentive structure. Such an experiment would also risk loss of control, as the experimenter cannot realistically maintain the control of a laboratory environment over several days. Thus, we direct the current study to the more subtle effects of *short-term* incentives in sustaining *long-term*

creativity, capturing the incremental effects of a longer time horizon while maintaining the same type of short-term incentives used in prior experiments. Our primary finding is that, relative to fixed pay, short-term quantity-based incentives have no measurable effect on high-creativity production at the time of the first-stage experiment, but such an effect emerges ten days later when we invite participants to complete the second stage.

In accounting, Kachelmeier, Reichert, and Williamson (2008) (hereafter, KRW) extend the experimental literature on incentives and creativity to a more production-like setting involving the design of "rebus puzzles," in which the firm (as proxied by the experimenter) values both the *quantity* of production and the *creativity* of that production. They find that quantity-based incentives increase participants' total production quantity without lowering the production of high-creativity ideas, as proxied in their study by creativity ratings in the overall top quartile. In this sense, quantity incentives also maximize their measure of "creativity-weighted productivity," insofar as all ideas contribute positively to this score even if rated relatively low for creativity. Importantly, however, neither KRW nor a follow-up self-selection study by Kachelmeier and Williamson (2010) find that quantity incentives actually *improve* creativity. Rather, the only gains they detect from quantity-based incentives come from increased total production, not the creativity of that production. We direct the current study to the deeper question of whether quantity-based incentives can go beyond improving production quantity to also aid in sustaining long-term *creativity* on the task to which the incentives apply.

Specifically, while quantity-based incentives may not improve creativity *initially*, our study examines the premise that an incentive to produce more ideas (of any creativity) serves to "prime the pump" with initial progress on the assigned task that can pay dividends later (Amabile and Kramer 2010; 2011). We test this premise by inviting participants to submit one or

more creative ideas for the same basic task, entirely at the participants' discretion, ten days after conducting an experiment similar to that of KRW. Participants are unaware of this invitation at the time of the first-stage experiment, and the second-stage task does not provide any incremental incentive other than the same offer of $10 additional fixed pay that we provide to all participants. That is, participants paid for quantity in the first-stage task do not earn anything more in the second stage that is not also given to their counterparts with fixed pay in the first stage, and the additional compensation of $10 for participation in the second stage does not depend on either the quantity or the creativity of participants' second-stage ideas.

Similar to prior research, we find a quantity effect of quantity-based incentives in the first-stage experiment, but we do not find a creativity effect in terms of the number of ideas that meet a high-creativity threshold. In the second stage ten days later, however, we find both a quantity *and a creativity* effect of quantity-based incentives provided in the first stage, even though those incentives no longer apply. The likely theoretical explanation is that creativity requires a "spark" of initial progress that motivates sustained attention over time. Our first-stage task likely provides this spark, motivating quantity-compensated participants to generate more ideas initially, even if those ideas are not any more creative than those provided by participants with fixed pay in the first stage. Our second-stage task then captures the extent to which the greater production achieved by quantity-compensated participants in the first stage eventually stimulates greater *creativity* as well. Providing supplemental support for this reasoning, our quantity-based participants self-report spending more time thinking about the creative task (i.e., designing rebus puzzles) in the ten days between the first and second stages, even without any monetary incentive to do so.

Our results suggest important implications for the use of performance-based pay in organizations that benefit from creativity. To our knowledge, ours is the first study to show that quantity-based performance incentives can further the goal of sustaining *high-creativity* production, even after the incentives have been removed. This finding helps to explain the popularity of performance-based incentive schemes in creativity-dependent firms (Grabner 2014). Second, our results stand in contrast to the perspective often advanced in the psychology and management literatures that extrinsic incentives undermine intrinsic motivation for creative tasks. Even if this phenomenon can be demonstrated in the experimental laboratory in the short term, our study suggests that a longer-term focus is essential to a more complete picture.

We review the literature and develop hypotheses in the next section, followed by describing the experimental method and design in Section III, presenting results in Section IV, and concluding in Section V.

II. Literature Review, Theory, and Hypotheses

The Effect of Performance-Contingent Incentives on Creativity

The question of how incentives influence creativity has long been of interest in psychology and management (Byron and Khazanchi 2012), and has recently been extended to management accounting settings, including experiments by KRW (2008), Kachelmeier and Williamson (2010), Chen, Williamson, and Zhou (2012), and a recent survey-based study by Grabner (2014).[1] A common theme from accounting studies is that, contrary to the view often expressed in psychology and management, creativity and performance-contingent incentives can be compatible. KRW find that a simple quantity-based incentive scheme outperforms fixed pay

[1] Our experiment is closest in design to KRW. Kachelmeier and Williamson (2010) extend KRW to the *self-selection* effects of creativity incentives, finding that more creative participants prefer creativity as an element of their compensation package and are more creative initially, although quantity-based pay eventually predominates in terms of creativity-weighted productivity. Chen et al. (2012) examine incentives for *group* creativity, finding that creativity incentives work best when implemented across groups, not within groups.

in terms of both quantity and creativity-weighted productivity, albeit with no better (and no worse) performance in terms of *high-creativity* output. Grabner's (2014) survey-based evidence finds that creativity-intensive firms tend to utilize a greater degree of performance-contingent incentives than do firms that depend less on creativity, although her study is not designed to measure the extent to which incentives actually improve creativity.[2]

While these studies show that performance-based pay and creativity can be compatible, they stop short of demonstrating that performance incentives can actually *enhance* creativity, particularly at the highest levels of creativity. More specifically, while KRW find that a quantity-only pay scheme outperforms all other pay schemes they test in terms of "creativity-weighted productivity," it is important to note that KRW's measure of creativity-weighted productivity is linked to total quantity by construction. That is, even if one weights workers' production quantity by creativity scores, the more one produces of *any* creativity, the higher is the creativity-weighted productivity score. As the authors explain, the creativity-weighted productivity advantage enjoyed by KRW's quantity-only participants is mostly a reflection of the "volume strategy" achieved from these participants' greater overall production. In contrast, for *high-creativity* production, which KRW define as ideas in the top quartile of the overall distribution of creativity ratings, quantity-only incentives result in a "dead heat" with the other incentive conditions they test (including fixed pay), with no discernable treatment differences. Essentially, KRW find that quantity-based pay increases overall productivity without increasing or lowering the number of high-creativity ideas.

Given that business success often depends on highly creative ideas (Fairbank and Williams 2001; Fallon and Senn 2006), we direct this study to the open question of whether

[2] Grabner's (2014) survey also does not specify the nature of the measure(s) on which performance-contingent incentives are based, such that our focus on quantity-based incentives is consistent with her setting and findings.

performance-based incentives can improve not just productivity in general, but *high-creativity* productivity in particular. We posit that KRW and similar studies did not find such an effect because their results are restricted to the short-term effects of incentives in a 20-minute task completed at the time those incentives are introduced. In the neurobehavioral literature, Dietrich (2004) provides evidence from brain research that sustained creativity requires "defocused attention" that transitions between the conscious and subconscious mind over time. When appropriately simulated, a defocused state of the brain works in the background, generating novel combinations of information that can yield creative insights (Dietrich 2004, 1016). This perspective helps to explain why incentive-based differences in creativity experiments are unlikely to arise instantaneously, which we address by designing a longer-term test.

More generally, a longer-term perspective on creativity incentives builds on the "progress principle" advanced by Amabile and Kramer (2010; 2011), which holds that long-term success in thought-intensive tasks arises in part from making meaningful short-term progress. In short, "priming the pump" with some initial progress reinforces the motivation to continue, which in turn leads to more progress. Amabile and Kramer do not advocate the use of extrinsic rewards for creative tasks, and in other writings, Amabile (1996) adopts the view common to social psychology that extrinsic incentives can undermine the intrinsic desire to be creative. Nevertheless, some elements of Amabile and Kramer's (2011) "progress principle" suggest the possibility that incentives can spark the initial progress that leads to long-term creative success. For example, in a study of workers who kept diaries of one significant event each day, Amabile and Kramer (2011) find that external forces such as deadlines led to significant progress, even if stressful at the time, which then improved the workers' subsequent success. Consistent with this

premise, Grabner (2014) finds from survey-based evidence that creativity-intensive firms employ *more* performance-contingent rewards than do firms that engage in more routine tasks.

Extended to a creative-design task, the progress principle suggests that quantity-based incentives could "prime the pump" with additional initial ideas. While the creativity of these ideas is likely to vary, the increased volume provides more practice with the task. Such practice might not lead to greater creativity *initially* because it is difficult to force creative ideas on the spot (KRW 2008), in addition to the fatigue factor of immediate incentives noted by Prendergast (1999) and the absence of the sustained, defocused attention that nurtures creativity in the brain (Dietrich 2004). Over time, however, if initial exposure to quantity-based pay simulates a greater immersion in the task, that immersion is likely to pay off later in terms of sustaining greater long-term creativity than would be exhibited by individuals receiving fixed pay. An interpretation consistent with this reasoning is that the initial increase in quantity motivated by quantity-based incentives helps people climb the learning curve faster for any particular creative task,[3] even if that advantage takes time to realize.

This argument notwithstanding, we acknowledge that the role of explicit incentives on creativity is controversial, with multiple authors taking the position that extrinsic incentives undermine intrinsic motivation and hence demotivate creativity and other thought-intensive activities (e.g., Deci 1971; Amabile 1996; Deci, Koestner, and Ryan 1999). Although the experiments cited in support of this conclusion are generally short-term in nature, the application of their arguments to a longer-term setting is ultimately an empirical question. Because the literature is characterized by competing arguments, we test two null hypotheses, one for initial creativity and the other for longer-term creativity. That is, while we are not predicting the null,

[3] The argument is not that quantity-based incentives improve creativity in general, but rather that such incentives could have positive spillover effects on the creativity of the particular task to which the incentives apply.

we test the null against two-sided alternative theory-based arguments that performance-based incentives could either improve or worsen high-creativity production.

H1: Relative to fixed pay, quantity-based compensation will have no effect on high-creativity production at the time of the initial experiment.

H2: Relative to fixed pay, quantity-based compensation will have no effect on high-creativity production several days after the incentives are removed.

Incentivizing Creativity Directly

The arguments discussed thus far apply to quantity-based performance incentives, which reward productivity but do not explicitly reward creativity. It is natural to also ask what effect *creativity*-based incentives might have on creative production. Accordingly, as a supplemental test to shed additional insights on our primary hypotheses H1 and H2, we also test a condition with *high-creativity* incentives instead of quantity-only incentives. KRW and Kachelmeier and Williamson (2010) find that adding creativity weightings to the compensation scheme *lowers* creativity-weighted productivity relative to quantity-only pay, without increasing the number of high-creativity ideas. In explaining this finding, KRW offer the reasoning that creativity is more a product of inspiration than raw effort, limiting the effectiveness of incentives that fixate people on the need to be creative. From the perspective of the progress principle, it would seem that the most initial progress would come simply from trying as many ideas as possible, as quantity-only incentives would reward, such that we have no *ex ante* reason to believe that high-creativity incentives would outperform quantity-only incentives in the long-run. The issue is ultimately an empirical question that we address later in the paper.

III. Method, Task, and Design

Task

We recruited 79 student volunteers from undergraduate business classes for participation in a compensated laboratory experiment. To facilitate comparability with prior research, we patterned the experimental task after the "rebus puzzle" experiments conducted by KRW (2008) and by Kachelmeier and Williamson (2010). As illustrated in Figure 1 with examples from the current study, 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 task is creative because participants *design* rebus puzzles rather than solve them, although we asked participants to write the solution at the bottom of each puzzle in order to facilitate creativity ratings, discussed later. For a creative production task, designing rebus puzzles has the advantage of meaningful variation in both quantity and creativity, as evidenced in prior research. We provided participants with the same instructional examples as those reproduced in the appendix to KRW (2008, 368-372), informing them (truthfully) that an independent panel of creativity raters would evaluate their submitted puzzles afterwards for creativity, "where creativity refers to puzzles that are original, innovative, and clever" (quoted from the instructions).

In the first-stage experiment, participants used a stack of blank index cards to design and submit rebus puzzles for 20 minutes, placing each puzzle in an "output box" when finished. Although KRW informed their participants in general terms that the researchers value quantity and creativity, our instructions focus on the high-creativity end of the scale, informing participants that "we value the *number of high-creativity puzzles* you can construct" (emphasis in original). We then informed participants (truthfully) 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 common understanding that we wanted them to submit high-creativity puzzles. Using a different task, Kachelmeier, Thornock, and Williamson (2015) find that value statements of this nature can influence behavior, which for our study is consistent with focusing participants' attention on our primary dependent variable, the number of high-creativity puzzles submitted.

Design

Our experimental design involves two manipulated factors: compensation scheme and experimental stage. Our first treatment factor, manipulated between participants, is the compensation scheme that participants face in the first stage, which either provides fixed pay of $25 as a control condition or performance-based pay that depends on the quantity produced. In the *fixed-pay* condition, participants (n = 26) read the following:

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

In contrast, participants in the *quantity-pay* condition (n = 27) read the following:

> Your compensation will be based on **how many** rebus puzzles you can construct in 20 minutes, irrespective of the creativity ratings of those puzzles. That is, to determine your compensation, we will simply ***count the number of rebus puzzles you submit, no matter what creativity ratings those puzzles receive***. For the participants using this version of the research, we will determine a cash payment rate that results in **$45.00** total compensation for the participant (or participants, if tied) with the ***highest*** number of puzzles submitted, and **$5.00** for the participant (or participants, if tied) with the ***lowest*** number of puzzles submitted. Everyone else in this version of the research will receive something in between $5.00 and $45.00, depending on the number of puzzles submitted, thus resulting in an expected average compensation around **$25.00**. In other words, the more puzzles you submit, irrespective of their creativity ratings, the more money you will

> make. We will pay you in two weeks, after we have analyzed the results to determine the payment rate that achieves this compensation.

By determining the pay rate afterwards, we are able to set the *average* compensation in the quantity-pay condition at $25, the same payment provided to all participants in the fixed-pay condition. Thus, our compensation scheme factor manipulates the *nature* of compensation without manipulating the average *amount* of compensation. This control is important because we are interested in the incentive effects of performance-based pay as a compensation design scheme, not merely the behavioral effects of paying more or less money to experimental participants.

As mentioned above, we also implement a third condition involving a *high-creativity* pay scheme (n = 26), operationalized in a manner similar to our quantity-pay scheme, but rewarding only high-creativity puzzles (i.e., those rated six or better). We comment on this this condition as part of our supplemental analyses.

Our second experimental factor, manipulated within participants, is the *experimental stage*. The first stage is the 20-minute experimental task described above, which we use as a within-participants control for comparison to high-creativity production in the second-stage task, discussed next. Importantly, participants do not know at the time of the first-stage experiment that there will be a second stage. We simply ask participants to come to a specified location ten days later within a specified block of time to collect their compensation for the experiment. Thus, participants have no reason to continue to think about the task between the two stages other than any motivation they might have developed from the initial experiment.

An important clarification is that our experiment does not implement *long-term* incentives, such as informing incentivized participants at the time of the first-stage experiment that we would continue to apply the same incentive structure for any additional ideas provided

ten days later. Although explicit long-term incentives might have led to a stronger long-term effect, the risk of such an approach is that, outside the control of the laboratory environment, participants might have been tempted to "cheat" by accessing rebus puzzles from online sources and/or from other participants. We desired an environment in which creative ideas emerge spontaneously from the participants' ideas, not from their research or extraneous efforts. Also, from a theoretical perspective, we wanted to test the extent to which the *initial* incentive prompts the sustained, defocused attention over time that Dietrich (2004) asserts is pivotal to creativity, apart from any behavioral influence of changing the incentive scheme itself. Accordingly, we took the conservative approach and designed the experiment to eliminate any incremental effect of explicit long-term incentives, while recognizing that long-term incentives could potentially generate a stronger effect.

Upon arrival at the specified room ten days after the initial experiment, we provided *all* participants with a document containing the following wording:

> 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.
>
> Presuming that you are willing to give us these few additional minutes before collecting your payment, please provide responses to the items on the next page. After you are finished, please hand these responses to the administrator, who will give you your cash compensation, including the additional $10. If you have to leave now and cannot give us a few more minutes, please see the administrator to get paid what you earned from last week's participation, but we certainly hope that you can spare just a few more minutes for an additional $10 of compensation.

As this wording indicates, we asked for the additional participation *before* providing participants with their first-stage compensation, thereby minimizing the risk of influencing quantity-pay participants in the second stage by revealing their earnings from the first stage. We also gave *all*

participants the same offer of $10 additional, fixed compensation for a few more minutes of time, such that there is no incremental treatment manipulation of participant incentives in the second stage. As previously discussed, the entirety of our between-participants incentive manipulation occurred ten days before we collected the second-stage responses.

The $10 incremental fixed pay offered to all participants for the second-stage task is important from a logistical and ethical perspective, avoiding any sense of experimenter deception from announcing a second-stage task of which participants were previously unaware. That is, participants across conditions were of the understanding that their pay for the first-stage task would be based entirely on the compensation promised for the first-stage (either fixed pay or performance-based), with no need to do anything else. To be consistent with this understanding and still gain the participants' willingness to complete a second-stage task at the time of collecting their first-stage pay, we offered an additional stipend as fair compensation for a separate request, which participants could accept or reject as they wished. Although participants could have turned down the offer and still collected their first-stage compensation as promised, all participants accepted the offer of $10 additional fixed pay for the few minutes required for the second-stage. The 100 percent acceptance rate stands to reason, given that participants could earn this $10 for submitting as little as one additional rebus puzzle, as we explain next.

Upon accepting the offer for $10 additional fixed pay, all participants read the following:

> We are interested in your approach to the task now that several days have elapsed since the experiment. In your envelope are a few more blank index cards. Please use ***one*** of these cards to construct the most creative puzzle ***that now comes to mind,*** writing the solution at the bottom of the card as you did before***. Please also write "MOST CREATIVE" on the back of the card.*** We are asking everyone to submit this one "most creative" puzzle. Then, if you wish to do so, at your option you can construct ***up to*** ten additional creative rebus puzzles, printing the solution to each puzzle at the bottom of the card. You will receive $10 additional compensation no matter how many additional puzzles you submit. When

completed, please place your puzzles in the envelope that contained these instructions.

Note from these instructions that our second-stage request was modest, asking for only one "most creative" puzzle from each participant to ensure that each participant would provide at least one data point. Participants then had the option to use up to ten additional blank index cards to provide up to ten more creative puzzles, but only if they wished to do so. Participants received $10 additional fixed pay irrespective of how many additional puzzle(s) they submitted. By asking participants to put any new puzzles in an envelope along with any unused index cards, we ensured that all submitted envelopes would be of the same thickness, thus minimizing any sense of experimenter pressure to design more puzzles. In short, our second-stage instructions attempt to capture the extent of participants' voluntary willingness to continue with the task from ten days earlier, but without providing any additional performance-based incentives.

The second-stage materials closed with a post-experimental questionnaire, which participants submitted along with the envelope containing any new puzzles. Participants then collected their cash compensation from the first-stage experiment and the $10 additional fixed pay for the second stage.

Creativity Ratings

To construct our dependent variable, we obtained creativity ratings from four panels of eight independent raters each – three panels for the first-stage experiment and a fourth panel for the second-stage puzzles. We recruited creativity raters from undergraduate business honors classes that did not provide experimental participants, thus avoiding any overlap between the participant pool and the rater pool. Raters received fixed compensation of $50 each for a rating session of about 2½ hours. Creativity raters first read the same background instructions and examples as did the experimental participants, but without any information on treatment

manipulations. They then used radio-frequency response devices to rate each puzzle on a 1 (lowest) to 10 (highest) scale, with puzzles projected one at a time. We needed four panels to mitigate rater fatigue, as each panel had to evaluate approximately 500 rebus puzzles. The order of puzzles was randomized for rating purposes, and raters were blind to treatment conditions. To ensure consistent initial calibration across panels, we began each rating session with the same set of 40 puzzles. For the fourth session of 230 second-stage puzzles, we added 170 first-stage puzzles (randomly selected and interspersed) to make the distribution of puzzles as similar as possible across rating sessions.

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

IV. Results

Descriptive Statistics and Overview of Findings

Table 1 shows descriptive statistics for the *overall* quantity and *high-creativity* quantity of rebus puzzles submitted in both experimental stages, with "high-creativity quantity" defined as puzzles receiving a creativity rating of six or better, consistent with the definition provided in the instructions. For the first-stage experiment, participants compensated for quantity produce significantly more puzzles (average of 24.15) than do those with fixed pay (average of 13.50), with the difference significant at $p < 0.001$. This difference is not surprising, given the incentive scheme, and is consistent with the similar result reported by KRW. The production advantage gained by quantity-compensated participants does not extend to high-creativity productivity in the first stage, however, in which the number of first-stage high-creativity puzzles submitted is 2.44 and 2.73 in the quantity-based pay and fixed-pay conditions, respectively. Overall, we

observe an average of 18.92 total puzzles and 2.58 high-creativity puzzles across conditions in the first stage, such that high-creativity production is 13.6 percent of total production, similar to the 15 percent estimate stated in the instructions.

Thus, for the first-stage experiment, participants operating under quantity-based pay submit significantly more puzzles in total, but about the same number of high-creativity puzzles as their fixed-pay counterparts. These first-stage findings are consistent with results reported previously by KRW. In the second-stage ten days later, participants submit only 2.34 puzzles in total, on average, which is far fewer than the average of 18.92 in the first stage. This large difference reflects the structural difference in how we implemented each stage. Specifically, the first-stage task asks participants to design rebus puzzles for a full 20 minutes. Participants come to the second-stage location to collect their first-stage compensation, at which time we offer $10 additional fixed compensation for as little as *one* additional puzzle. Participants then have the option to provide up to ten more second-stage puzzles, but only if they wish to do so. Predictably, many participants provided one second-stage puzzle, collected their compensation, and left. As we report next, however, we do observe a difference from the first-stage treatment manipulation in the extent of participants' second-stage high-creativity production.

Participants compensated for quantity in the first stage continue to submit more puzzles *in total* than their fixed-pay counterparts in the second stage (2.74 vs. 1.92), but the more striking observation is that almost all of this difference is due to puzzles that meet the high-creativity threshold. Specifically, 1.44 of the 2.74 second-stage puzzles submitted by participants assigned to the quantity-based pay condition in the first stage reach the high-creativity threshold, on average, in comparison to only 0.69 of the 1.92 second-stage puzzles submitted by those assigned previously to fixed pay.

Tests of Hypotheses

To test our hypotheses, we standardize the dependent variable by computing Z-score transformations of high-creativity production within each stage of the experiment, subtracting the overall mean high-creativity production for each stage from each corresponding observation and dividing by the overall standard deviation for each stage. After this transformation, the overall mean within each stage is zero by construction, thereby controlling for the structural difference between the two experimental stages. Nevertheless, we can still test the compensation scheme × stage *interaction* to examine whether our primary manipulation of quantity-based vs. fixed pay has different effects across the two experimental stages.

We add a covariate to the analysis for *openness*, extracted from one of the personality dimensions in the "Big-Five Inventory Questionnaire" (John, Donahue, and Kentle 1991). McCrae (1987) finds a significant correlation between individual creativity and tolerance for openness, which captures "an interest in varied experience for its own sake" (1987, 1259). In short, more "open" people are more receptive to new and different things. Finally, of the 240 second-stage puzzles submitted, ten are duplicates of the same participants' first-stage puzzles. Submitting a duplicate is inconsistent with our research objective of assessing differences in second-stage creativity, so we delete these ten observations from all second-stage analyses.[4]

An analysis of covariance (ANCOVA) on Z-score-transformed high-creativity production with the openness covariate generates a significant compensation scheme × stage interaction ($F = 5.05$; two-tailed $p = 0.029$), as depicted in Figure 2. As this figure shows, after standardizing via the Z-score transformation to remove the overall decline in the quantity of puzzles submitted from the first stage to the second stage, quantity-based pay appears to benefit high-creativity production only in the second stage of the experiment, even though the

[4] Our research conclusions do not depend on deleting the ten second-stage duplicates.

compensation scheme itself applies only to the first stage. Below, we report statistical tests for each of our hypotheses.

For H1, Panel A of Table 2 reports no statistically significant differences in high-creativity production at the time of the initial experiment attributable to quantity-based pay or the openness covariate. Thus, we fail to reject null hypothesis H1. This finding is consistent with prior research by KRW, and likely reflects Dietrich's (2004) neuroscientific perspective that creativity does not happen instantly. Thus, although participants paid for quantity respond to this incentive by submitting significantly more first-stage puzzles do than their fixed-pay counterparts, they are unable to transform this initial advantage into a greater number of first-stage *high-creativity* puzzles.

For H2, we obtain a different result. As Figure 2 depicts and Panel B of Table 2 tabulates, participants paid for quantity in the first stage of the experiment generate significantly more high-creativity puzzles *in the second stage* than do their fixed-pay counterparts ($F = 6.53$; two-tailed $p = 0.014$), even though the quantity-based incentive no longer applies in the second stage. Thus, we reject null hypothesis H2 in favor of the alternative hypothesis that, relative to fixed pay, quantity-based pay improves high-creativity production long after the incentive is removed. We also see a significant second-stage effect for openness ($F = 5.79$; $p = 0.014$).

Our findings for H2 pose a challenge to the view often expressed in psychology and management that incentives undermine creativity. However, our findings are consistent with the "progress principle" (Amabile and Kramer 2010; 2011) that meaningful experience towards a goal, even if that goal is not attained initially, helps to further long-term success. In our study, quantity-compensated participants generated significantly more puzzles than did fixed-pay participants in the first stage of the experiment. These additional puzzles *did not* translate to

additional high-creativity puzzles in the first stage, but they likely served to "prime the pump" with additional meaningful experience that influenced participants after the experiment ended. Ten days later, participants initially compensated for quantity voluntarily submitted not just a higher number of second-stage puzzles, but also puzzles that were more creative than those submitted by their fixed-pay counterparts. This finding goes beyond KRW and related experiments to show that performance-based incentives can yield both short-term quantity benefits and long-term *creativity* benefits.

Supplemental Analyses

Alternative Measure of Second-Stage High-Creativity Production

Recall that our second-stage instructional materials request only one "most creative" puzzle, with the option to submit up to ten more puzzles "if you wish to do so." Predictably, given these instructions, the modal participant in stage two submitted only one puzzle, which sometimes met the high-creativity threshold and sometimes not. This structure suggests a natural division of our second-stage responses into three categories: (1) those who submitted no high-creativity puzzles, (2) those who submitted one high-creativity puzzle, and (3) those who submitted two or more high-creativity puzzles. Coding these three categories as 0, 1, and 2, respectively, in an ordered logistic regression supports the conclusion of greater second-stage creativity among those compensated for quantity in the first stage (two-tailed $p = 0.046$, untabulated). Eight of the 27 participants (30 percent) in the quantity-based pay condition fall into the highest category of two or more second-stage high-creativity puzzles, in contrast to only two of the 26 participants (8 percent) in the fixed-pay condition. Importantly, if we apply the same categorical test in the first-stage experiment, we *do not* observe a significant difference

(p = 0.205, in the opposite direction), supporting the conclusion that performance-based incentives can help to *sustain* creativity even if they have no effect on initial creativity.

Change in Creativity

Because the second-stage task does not have any set time period and only asks for (at least) one puzzle, it is not surprising that participants submit far fewer puzzles in the second stage than in the first. Accordingly, from the simple aspect of total volume, the second stage offers fewer opportunities to reach the high-creativity threshold. Given this environment, it is interesting to examine the most creative puzzle submitted by each participant in each stage. For 52 percent (28 percent) of the participants in the quantity-compensated condition (fixed-pay condition), the most creative puzzle submitted in the second stage is rated at equal or greater creativity than the most-creative puzzle submitted in the first stage. These percentages significantly differ in a logistic analysis that controls for the openness covariate (two-tailed p = 0.05). This finding corroborates the conclusion that quantity-compensated participants were more successful than their fixed pay counterparts in sustaining high creativity from the first to the second stage.

Mediation Analysis

The "progress principle" that we draw upon as the theoretical basis for our findings rests on the notion that people have to experience meaningful progress towards a goal in order to achieve more success meeting that goal. For our task, we observe no difference in high-creativity production between quantity-compensated and fixed-pay participants in the first-stage experiment. Accordingly, the differential progress between the two incentive conditions likely arises from the increased first-stage production in the quantity-compensated condition of "reasonable" puzzles that were close to but did not quite reach the high-creativity threshold. To

capture this progress, we extract the number of "moderate-creativity" puzzles generated by each participant in the first-stage experiment, defining this category as puzzles attaining a rating greater than four but less than six. Reflecting their overall greater volume, quantity-compensated participants enjoy a significant first-stage advantage over their fixed-pay counterparts in moderate-creativity production ($p = 0.003$).

We then employ the first-stage moderate-creativity count in a bootstrapping-based mediation analysis, as recommended by MacKinnon et al. (2002) and explained procedurally by Preacher and Hayes (2004; 2008). Based on the results of 1,000 bootstrapped samples, this procedure supports the conclusion that first-stage quantity-based pay exerts an indirect effect on second-stage high-creativity production that is mediated by first-stage moderate-creativity production ($p < 0.05$). Importantly, we do not obtain the same conclusion from an alternative mediation variable based on the number of first-stage *low-creativity* puzzles (i.e., those rated lower than four), for which the bootstrapping analysis does not support a mediating effect. Thus, it would appear that "progress" for our task hinges on experiencing some success in the first stage that at least gets close to the high-creativity threshold. With this progress, participants are more likely to sustain high-creativity production after the incentives have been removed. Because quantity-compensated participants achieve more first-stage progress at the moderate-creativity level, they sustain more high-creativity success in the second stage.

Time Spent Thinking About the Task Between the First and Second Stages

If creativity requires the nurturing of "defocused attention" in the brain over time (Dietrich 2004), we should see some evidence of this differential attention in terms of time spent thinking about the task between the first and second stages of the experiment. To shed some insight towards this end, we asked participants in a post-experimental questionnaire administered

after the second-stage task to self-report the number of minutes spent thinking about rebus puzzles over the days after the first-stage experiment. These self-reports exhibit high variance, and are at best a noisy proxy for the actual time spent. Nevertheless, we find at least marginally significant evidence that participants with quantity-based pay in the first stage self-report spending more time thinking about rebus puzzles in the days following the first-stage experiment than do participants with fixed pay in the first stage (two-tailed $p = 0.085$, using a rank-transformed measure of time spent to reduce the variance of responses).

We do not wish to overinterpret this finding, as we are unable to find significant evidence that self-reported time spent mediates our primary findings, nor do we even find systematic evidence that self-reported time spent thinking about rebus puzzles after the first-stage experiment is correlated with second-stage high-creativity production.[5] As noted, a self-reported measure of time spent from a post-experimental questionnaire is a noisy proxy at best, and accordingly does not lend itself well to rigorous follow-up analyses. At a minimum, however, we can say that self-reported time spent is consistent with our theoretical premise that initial quantity-based incentives "prime the pump" with initial progress that stimulates greater attentiveness to the task objective (in our case, coming up with creative rebus puzzles) over time.

Supplemental Treatment Condition: High-Creativity Incentives

Insofar as our primary dependent variable is high-creativity production, it is natural to question whether explicit incentives for high-creativity ideas would exhibit different effects from those we observe for quantity-only incentives. Intuition might suggest that the best way to motivate high creativity is to incentivize high creativity, but there is reason to expect otherwise, as KRW find that adding creativity to the incentive scheme fixates individuals too much on

[5] One exception is that, if we split the sample on self-reported task familiarity, we find a marginally significant positive correlation between self-reported time spent thinking about rebus puzzles and second-stage high-creativity puzzles among the subsample that falls below the median on familiarity with rebus puzzles (two-tailed $p = 0.069$).

being creative rather than being productive. Because it is difficult to force creativity, such fixation leads KRW's creativity-compensated workers to be less productive overall but no more successful in high-creativity production than those paid for quantity only. Extended to our interest in the progress principle, if participants compensated for high-creativity puzzles are less productive in the first stage than those compensated for overall quantity, they would experience less initial progress, and hence would have less to draw on for sustained creativity ten days later.

Consistent with this reasoning, we test a supplemental "high-creativity" treatment condition (n = 26) in which we compensate participants in a manner similar to the quantity-based condition in the first-stage experiment but reward only those puzzles rated six or higher. Participants in this condition produce a first-stage average total quantity of 14 that is nearly identical to that observed in the fixed-pay condition, while significantly lower than the first-stage quantity of 24 generated by participants in the quantity-only incentive condition. The number of first-stage high-creativity puzzles is nearly identical across all three conditions (average of 2.5). In the second stage, participants compensated for high-creativity puzzles in the first stage achieve second-stage high-creativity production that lies in between the quantity-based and fixed-pay conditions. Specifically, first-stage high-creativity-compensated participants produce an average of 0.92 high-creativity puzzles in the second stage. This average is more than the 0.69 produced in the second stage by participants with fixed pay, but it is less than the 1.44 produced by participants compensated for quantity only. Neither of these comparisons with the supplemental "high-creativity" incentive condition is statistically significant at conventional levels, although, as reported earlier, the fixed-pay and quantity-based incentive conditions significantly differ from each other.

On balance, it seems safe to conclude that, if the goal is sustained creativity over time, initial incentives for high-creativity production fare no better than incentives for total quantity. From the perspective of the progress principle, quantity-based compensation stimulates the greatest initial experience (in the sense of volume) from which workers can draw on later. It is interesting to contemplate this reasoning in the context of Grabner's (2014) recent survey-based evidence on the use of performance-based incentives among creativity-intensive firms in practice. Grabner finds that creative intensity leads firms to be *more* inclined to use performance-based compensation, although her survey does not specify the exact kind of performance that performance-based compensation is rewarding. We suggest as an avenue for future field-based research that it would be interesting to extend Grabner's (2014) findings to explore whether creativity-intensive firms are explicitly rewarding *creativity*, as opposed to rewarding other dimensions of productivity that generate more progress towards high-creativity production even if they do not reward creativity directly.

V. Conclusions

Management accountants have long been interested in the effects of performance-based incentives. Only recently, however, has this interest extended beyond routine production exercises to the softer dimensions of productive quality, such as creativity. Creativity is of interest not only because of its importance to contemporary business, but also because creativity does not necessarily arise from immediate effort. This characteristic poses a significant challenge to laboratory experimentation, as experiments have the advantage of controlled, *ceteris paribus* manipulation of treatment factors, but also carry the disadvantage of a short-term focus. We mitigate this disadvantage by collecting additional data on creativity ten days after a laboratory experiment. Our primary finding is that performance-based compensation that rewards quantity

in a first-stage experiment has no effect on high-creativity production at that time, relative to fixed pay. Ten days later, however, those paid for quantity in the initial experiment voluntarily submit significantly more ideas that meet a high-creativity rating threshold than do those with fixed pay. Indeed, our fixed-pay participants generally "gravitate to the bottom" in terms of providing minimal second-stage production and creativity. These findings highlight the importance of capturing creativity in longer-term settings.

Our findings pose a challenge to the view often expressed in the psychology and management literatures that explicit incentives are harmful for creative and other thought-intensive tasks. This claim often rests on short-term results from the experimental laboratory. In contrast, our longer-term perspective aligns more closely with a neurobehavioral perspective on creativity, which holds that creativity requires a sustained, defocused attention over time (Dietrich 2004). We test the potential for performance-based incentives to serve as the catalyst that seeds the initial progress necessary to motivate this longer-term sustained attention.

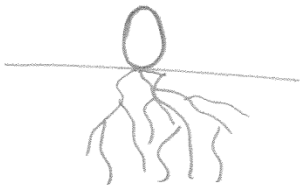
We encourage further research on the effects of incentive-compensation schemes on creativity and other performance dimensions that characterize an idea-based economy. Ideally, this research can draw from multiple methods, as is the case with recent experimental and field-based accounting research on creativity. What our study adds to the mix is a longer perspective than that captured in most laboratory experiments. That being said, our study is limited by the fact that our second-stage collection of data occurs in a setting in which participants have limited time and no advance knowledge of our request for additional participation. These features enable us to isolate the long-term effects of initial incentives that arise without any additional long-term incentives or anticipation of long-term requests. Future research can expand on our baseline by

incorporating explicit long-term incentive and feedback structures that facilitate learning and practice along with the basic incentive effects we identify.

Figure 1
Examples of Second-Stage Rebus Puzzles Submitted by Participants

Panel A: High-Creativity Puzzles

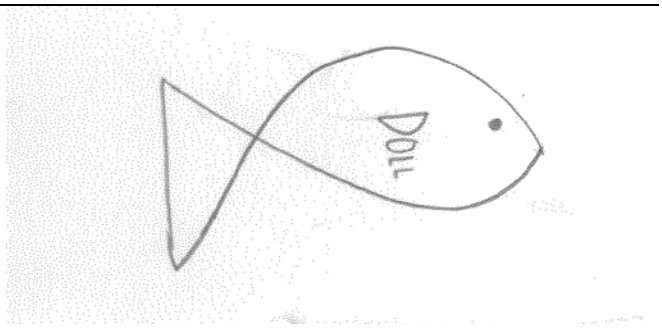
Solution: Obama
Creativity rating: 8.67
Condition: Fixed pay



Solution: Eggplant
Creativity rating: 7.83
Condition: Quantity-based pay

Panel B: Borderline High-Creativity Puzzles

Solution: Widespread panic
Creativity rating: 6.00
Condition: Fixed pay



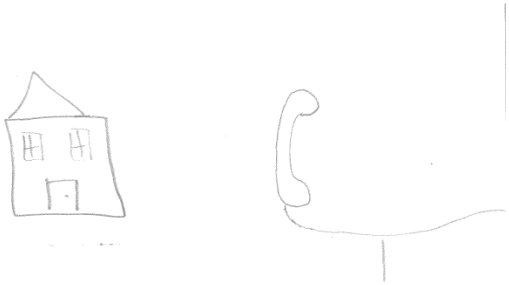
Solution: Dolphin
Creativity rating: 6.00
Condition: Quantity-based pay

Panel C: Moderate-Creativity Puzzles

Panel E

Solution: Cover-up
Creativity rating: 4.17
Condition: Fixed pay

Panel F



Solution: House Call
Creativity rating: 4.00
Condition: Quantity-based pay

Figure 2
Standardized High-Creativity Production by Compensation Scheme and Stage

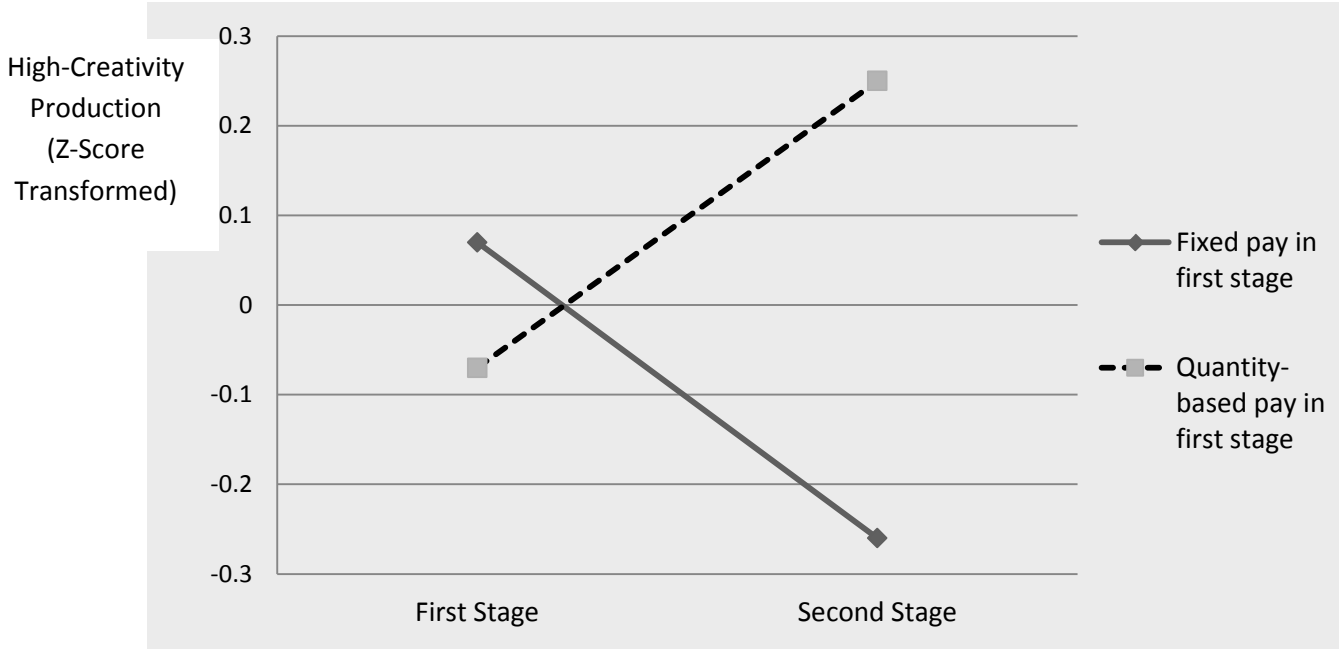


Table 1
Descriptive Statistics for Rebus-Puzzle Production: Means and (Standard Deviations)

	Fixed Pay in First Stage	Quantity-Based Pay in First Stage	Overall
Number of participants	26	27	53
Panel A: First-stage experiment			
Total quantity	13.50	24.15	18.92
	(4.93)	(12.21)	(10.73)
High-creativity puzzles	2.73	2.44	2.58
	(1.73)	(2.38)	(2.07)
Z-Score transformation of high-creativity puzzles[a]	0.07	-0.07	0.00
	(0.84)	(1.15)	(1.00)
Panel B: Second-stage follow-up			
Total quantity	1.92	2.74	2.34
	(2.30)	(2.90)	(2.63)
High-creativity puzzles	0.69	1.44	1.08
	(0.88)	(1.85)	(1.49)
Z-Score transformation of high-creativity puzzles[a]	-0.26	0.25	0.00
	(0.59)	(1.24)	(1.00)

[a] The Z-score transformation of high-creativity production adjusts for the structural differences between the two experimental stages by subtracting the overall mean and dividing by the overall standard deviation within each stage. Removing the main effect of stage in this manner and standardizing for the different variances across stages allows direct comparison of the compensation conditions across stages.

Table 2
Tests of Hypotheses

	df	MSE	*F*	*p*-value
Test of H1: First-stage high-creativity production (after Z-score transformation):				
Compensation scheme	1	0.18	0.18	0.675
Openness covariate	1	0.06	0.06	0.810
Error term	50	1.05		
Test of H2: Second-stage high-creativity production (after Z-score transformation):				
Compensation scheme	1	5.12	6.53	0.014
Openness covariate	1	5.79	6.67	0.012
Error term	50	0.87		

References

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

Amabile, T. M., and S. J. Kramer. 2010. What really motivates workers. *Harvard Business Review* 88 (1): 44-45.

Amabile, T. M., and S. J. Kramer. 2011. *The Progress Principle: Using Small Wins to Ignite Joy, Engagement, and Creativity at Work.* Boston, MA: Harvard Business Review Press.

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

Deci, E. L. 1971. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology* 18 (1): 105-115.

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.

Dietrich, A. 2004. The cognitive neuroscience of creativity. *Psychonomic Bulletin & Review* 11 (6): 1011-1026.

Fairbank, J. F., and S. D. Williams. 2001. Motivating creativity and enhancing innovation through employee suggetion 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.

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

Hannan, R. L. 2005. The combined effect of wages and firm profit on employee effort. *The Accounting Review* 80 (January): 167-188.

John, O. P., E. M. Donahue, and R. L. Kentle. 1991. *The Big Five Inventory – Versions 4a and 54.* Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

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., T. A. Thornock, and M. G. Williamson. 2015. Communicated values as informal controls: Gaining accuracy while undermining productivity? *Contemporary Accounting Research*, forthcoming.

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.

McCrae, R. R. 1987. Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology* 52(6): 1258-1265.

McGraw, K. 1978. The detrimental effects of reward on performance: a literature review and a prediction model. In *The Hidden Costs of Reward*, edited by M. Lepper and D. Greene. Hillsdale, NJ: Lawrence Erlbaum Associates.

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: 717-731.

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.

Prendergast, C. 1999. The provision of incentives in firms. *Journal of Economic Literature* 37(1): 7-63.