

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Stephen E. Deason
Stephen Deason

07/25/2017
Date

Incentivizing Creativity:
Cognitive Depletion and Multi-Dimension Incentive Contracts

By

Stephen Deason

Master of Business Studies

Accounting

Kristy Towry, Ph.D.
Advisor

[Advisor's signature]

Greg Waymire, Ph.D.
Advisor

[Advisor's signature]

Anand Swaminathan, Ph.D.
Committee Member

[Member's signature]

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Incentivizing Creativity:
Cognitive Depletion and Multi-Dimension Incentive Contracts

By

Stephen Deason
B.Sc. University of Alabama – Birmingham, 1993; M.B.A. Emory University, 2010

Advisors:

Kristy Towry, Ph.D.

Greg Waymire, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Business Studies
2017

Abstract

Incentivizing Creativity: Cognitive Depletion and Multi-Dimension Incentive Contracts

By Stephen Deason

In a recent influential study, Kachelmeier and Williamson (2010; KW) extend a series of experimental accounting papers investigating whether sorting contracts effectively sort employees into high- and low-creative types. KW provide evidence that the sorting value of contracts rewarding creativity is limited to the initial (i.e. short-run) production of creative ideas. This result seems to imply that firms interested in hiring high-creativity individuals would be better off - in terms of maximizing creative output - if they compensated employees using a one-dimensional contract rewarding quantity alone. Yet, the underlying premise of the balanced scorecard seems to be that multi-dimensional contracts improve realized outcomes. This study experimentally examines whether KW's findings are influenced by the depletion of an internal cognitive resource that may not impact these contracts in practice. Experimental results are opposite the findings of KW. I also find that operationalized cognitive resource replenishment using glucose appears to move worker productivity opposite the predicted direction. Further analyses demonstrate the importance of considering the underlying psychological and physiological mechanisms in research evaluating the effectiveness of multi-dimensional incentive contracts.

Incentivizing Creativity:
Cognitive Depletion and Multi-Dimension Incentive Contracts

By

Stephen Deason
B.Sc. University of Alabama – Birmingham, 1993; M.B.A. Emory University, 2010

Advisors:

Kristy Towry, Ph.D.

Greg Waymire, Ph.D.

A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of
Business Studies
2017

Table of Contents

I. Introduction	1
II. Background, Theory and Hypothesis	3
III. Method	10
IV. Results	13
a. Contract Selection Benefits	14
b. Hypothesis Tests	16
V. Conclusions	20
VI. References	24
VII. Figures	26
a. Figure 1	26
b. Figure 2	27
c. Figure 3	28
d. Figure 4	29
e. Figure 5	30
VIII. Tables - Descriptions	31
a. Table 1	32
b. Table 2	35

Incentivizing Creativity: Multi-Dimensional Incentive Contracts & Cognitive Depletion

I. INTRODUCTION

Firm success often depends on high productivity and creative innovation (Chang and Birkett 2004), yet whether and how to best incentivize creativity remains an open question. An accounting answer suggested by Kaplan and Norton (1992, 1996) encourages firms to measure and reward both efficient production and creative effort as inputs to strategic performance management systems (e.g. the Balanced Scorecard). Extending this concept, a pair of recent journal articles investigates whether pay-for-performance contracts rewarding creativity induce greater creative output (Kachelmeier, Reichert & Williamson 2008; henceforth KRW), and whether creativity-incentivizing contracts are effective *sorting* mechanisms (Kachelmeier and Williamson 2010; henceforth KW) in that contracts rewarding creativity may benefit firms by acting as “sorting contracts” that attract more creative workers (Demski & Feltham 1978; Sprinkle & Williamson 2007).

KW use a rebus puzzle task to examine whether performance-based compensation schemes effectively attract and motivate workers whose abilities align with the terms of the contract. In a rebus puzzle task, “participants design puzzles that use words, symbols and pictures to represent a familiar term or phrase” (KW, p. 1670).¹ Results from KW indicate “sorting contracts” effectively attract and incentivize creativity, but work only for *initial* creative production. KW’s counter-intuitive findings include that participants who *self-select* into a contract rewarding quantity alone (quantity-only condition) produce significantly more puzzles overall, and as just as many high-creativity puzzles, as participants who self-select into a contract rewarding both quantity and creativity (creativity-weighted condition).² KW interpret their results as supporting the value of a sorting contract for *initial* productivity, but not for long-run creative output.

This paper presents theory and experimentally tests whether KW’s results are attributable to self-selected participants in the creativity-weighted condition depleting the cognitive resource responsible for

¹ Please see “Supplemental Instrument: Experimental Instructions” for examples of rebus puzzles.

² KW define high-creativity as puzzles ranking in the top quartile of all puzzles produced; this study follows their definition.

coordinating thoughts and actions into purposeful behavior (Miller & Wallis 2009). Prior research indicates the most likely depleted resource is Working Memory (WM; e.g. Baumeister et al. 1998; Muraven & Baumeister 2000). The depletion of WM is important because WM directs and controls attention (Baddeley & Hitch 1974). Indeed, recent studies demonstrate that “sufficient ‘free’ WM is needed to perform well on creative tasks” (De Dreu et al. 2012, p. 659).

This study is important because academic research into rewarding creativity has failed to produce agreement regarding the effect of rewards on creative output (Eisenberger & Shanock 2003 provide a review). Although recent influential accounting studies (i.e. KRW and KW) provide some insight into whether firms can effectively reward creativity, KW’s results imply that practitioners cannot effectively use incentive contracts to sort long-term employees into more vs. less creative types. This result is unsatisfying because creative innovation is a dimension of performance known to drive organizational success (Chang & Birkett 2004); therefore, incentivizing creativity is important to audit firm partners (Chang & Birkett 2004), general business practitioners, and accounting researchers (e.g. KRW, KW). The present study tests whether KW’s findings result from the experimental environment and hence are not generalizable to practice.

In order to address whether KWs results are due to cognitive depletion, I perform an experiment that compares a replication of the self-selection portion of KW’s experiment to a condition that attempts replenishment of the depleted cognitive resource, observing the effects of WM replenishment on creative production. Specifically, I implement and measure results under the two incentive contracts considered by KW: (1) a piece-rate contract rewarding only quantity (quantity-only), and (2) a piece-rate contract rewarding quantity weighted by creativity ratings (creativity-weighted). Because participants self-select their contract type, this variable is measured, not manipulated. I randomly assign participants into two treatments: (1) cognitive replenishment, and (2) placebo. Participants are blind to which treatment they receive, and the main dependent variable is creativity-weighted output operationalized using three measures: (1) aggregate creative output, (2) high-creativity output, and (3) mean creativity ratings.

My results do not replicate the findings of KW. Specifically, compared to participants in the quantity-only condition, participants in the creativity-weighted condition are equally productive in terms of the cumulative number of puzzles produced per person while also producing both (1) a significantly greater number of high-creativity puzzles (i.e. puzzles ranked in the top quartile), and (2) significantly greater creativity ratings. These results are the reverse of KW's findings, and I find that cognitive replenishment actually *reduces* creative output, which is the opposite of the theory-supported result I predict. Supplemental analyses (discussed in Section IV) suggest a few reasons for the mismatch between my results, my predictions and KW's results.

The remainder of the paper is organized as follows: Section II provides background and develops theory-based hypotheses; Section III explains the method and design; Section IV presents results, and Section V concludes.

II. BACKGROUND, THEORY & HYPOTHESES

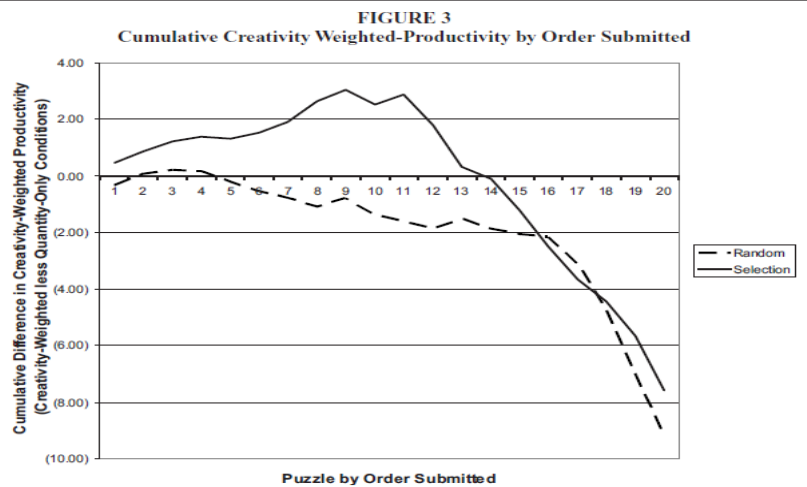
Academic accounting and practitioner interest in multi-dimensional performance-based compensation schemes continue to increase. For example, a recent study by Chang and Birkett (2004) finds that professional service firm survival relies on both high productivity and creative innovation. Yet whether creativity incentives increase (1) overall creative output or (2) the quality of creative output remains an unanswered question - as is the related question of how to properly implement multi-dimensional performance-based incentive schemes.

KW is among the first experimental studies investigating contract selection benefits in a *multi-dimensional* setting. Specifically, KW examine whether the contract-selection benefits of rewarding both quantity and creativity outweigh the divided effort costs (also see e.g. Christ et al. 2012) of a multi-dimensional compensation scheme.³ KW measure creative output in terms of three dependent variables: (1) aggregate creative output, (2) number of high-creativity puzzles, and (3) mean creativity ratings. Using the same “rebus” puzzle task from KRW, KW find that, when contracts are randomly assigned, participants in the quantity-only contract generate higher cumulative creativity-weighted productivity than

³ One possible hypothesis is that the effort-influencing “costs” of attending to multiple performance dimensions detracts from achieving both goals, resulting in lower overall output across both dimensions.

participants in the creativity-weighted contract. This paper investigates whether the difference KW find between the cumulative creativity-weighted and quantity-only conditions occurs because a limited cognitive resource is being depleted for participants in the creativity-weighted condition.

KW find that participants who self-select into the creativity-weighted contract are *initially* more productive, but their cumulative creative output is less than that of participants who self-select into the quantity-only contract *in the long-run*. The following graphic, taken from Kachelmeier and Williamson (2010, page 1682), is helpful in describing these effects. Each line in the graph represents the difference in cumulative creativity-weighted production for each puzzle produced; that is, the lines graph the cumulative creativity-weighted production of participants working under the creativity-weighted contract *minus* the cumulative creativity-weighted production of participants working under the quantity-only contract for each of the Random and Self-Selection treatments. Because this study is primarily concerned with the sorting contract, the self-selection condition (represented by the solid line) is most interesting. In particular, the solid line seems to provide information that the effectiveness of the sorting contract is subject to diminishing marginal returns for participants selecting the creativity-weighted contract.



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

One possible conclusion - the one drawn in KW - is that contract selection benefits are good for *initial* production; but, they are less useful for creative productivity in the *long-run*. In this paper, I propose that such an argument, along with its associated inferences, fails to account for the physiological mechanisms and chemical changes that cause the usefulness of the sorting contract to decrease over time (and/or with the number of puzzles produced). The basic idea is that creativity is a high-intensity cognitive task, and focusing on creative output requires self-control which depletes an internal cognitive resource. It may be that the decline of creative output for KW's participants in the creativity-weighted contract condition is due to the depletion of this cognitive resource. The following paragraphs expand on these terms, present the underlying theory, and develop testable hypotheses based on psychological and neurophysiological research in creativity and creative output.

In particular, the remainder of this section develops theory and discusses the links between KW's study and cognitive resources which are vulnerable to depletion and which are also linked to creativity. These include the Central Executive (CE), Working Memory (WM) and the role played by self-control. I begin by briefly elaborating on academic research into the effect of rewards on creative output. I next describe the potentially depleted cognitive resources of interest and their importance to creativity. I close this section by tying these concepts into recent accounting studies and presenting testable hypotheses.

Amabile (1988 p. 126) defines creativity as the "production of novel and useful ideas." Academic research in psychology and creativity has failed to produce general agreement regarding the effects of rewards on creative output (Eisenberger & Shanock 2003). Two well-developed psychology research camps hold opposing views on the effects of rewarding creativity. One view asserts that creativity is intrinsically motivated and therefore reduced by external rewards (e.g. Amabile 1996; Deci & Ryan 1985; Deci, Koestner & Ryan 1999; Mumford 2003). The opposing view asserts that rewards offered for novel performance facilitate creativity and creative innovation (e.g. Winston & Baker 1985; Eisenberger, Pierce & Cameron 1999; Freidman 2009; Stokes 2012). Although rewards may decrease creative output because they draw attention to the reward itself and remove attention from the creative process (e.g.

Amabile 1983, 1996), accounting incentives exist to influence or facilitate employee and managerial decisions and actions (e.g. Demski & Feltham 1976; Farrell et al. 2008).

In order to influence and facilitate decisions and actions, incentives draw on cognitive processes such as self-control (i.e. inhibition of habituated responses), task persistence (mental and physical), and emotional regulation (see e.g. Baddeley 2003; Gailliot 2008; Miller & Cohen 2001). Yet use of these cognitive processes has been linked to the depletion of a pooled internal energy resource used by Working Memory (WM) and the Central Executive (CE; De Dreu et al. 2012).⁴ As part of WM, the CE is the overarching cognitive system responsible for coordinating thoughts and actions into purposeful behavior (Miller & Wallis 2009). Specifically, the role of the CE and WM is to enable self-control, task persistence, and emotional regulation (Baddeley 2003; Gailliot 2008; Miller & Cohen 2001). Recall the depletion of WM is particularly important because WM is the ability to direct and control attention (Baddeley & Hitch 1974).

“Working memory refers to cognitive processes that retain information in an unusually accessible state, suitable for carrying out any task with a mental component. The task may be language comprehension or production, problem solving, decision making, or other thought” (Cowan 1999 p. 62). In addition, research indicates creativity relies on WM; specifically, “...working memory is also where creativity and innovation are born” (Vandervert, Schimpf & Liu 2007, p. 3). WM components include not only the CE, but also two additional slave functions: the visuospatial sketchpad and the speech loop (e.g., Baddeley 1992; Baddeley & Logie 1999; Cowan 1999).

The important point here is that consistently engaging any of the WM or CE functions depletes an internal energy resource (e.g. Baumeister et al. 1998; Baumeister & Heatherton 2004; Muraven & Baumeister 2000). Equally importantly, Gailliot et al. (2007) show all WM / CE functions drain a pooled resource. This means engaging any of the WM / CE functions reduces the availability of related

⁴ Notably the psychology literature does not clearly distinguish between WM and the CE. On the one hand, the CE is often conceived of as a part of WM (and the most important one, at that); on the other hand, WM could not exist without the CE acting in its role to coordinate thought and direct attention. For this reason, I attempt to consistently use WM rather than CE in this paper; however, for the purpose of this paper, the concepts are effectively interchangeable. Logically, creativity will suffer in the absence of thought or when it is difficult to handle distractions (i.e. maintain focus).

resources such as WM and Self-Control. The implication from these experiments (e.g. Baumeister et al. 1998; Baumeister & Heatherton 2004; Muraven & Baumeister 2000; Gailliot et al. 2007) is that CE depletion rapidly induces physiological changes that are relevant to incentive-contract research (e.g. decreases in self-control and task persistence). Thus, seemingly unrelated behaviors can affect one another by drawing down a pooled (common) internal energy reserve (Miller 2012). One activity known to drain this common resource pool of internal energy is the exercise of self-control.

Acts such as consciously and actively resisting temptation, making choices and planning for the future are examples of exercises in self-control. Baumeister & Heatherton (1996) describe self-control as any attempt to alter, override or otherwise manage one's habitual, normal or natural response. Studies have demonstrated that people who are better at self-control enjoy greater success in school (Mischel et al. 1988) and at work (Tangney et al. 2004), have higher self-esteem, experience better health, and enjoy more satisfying interpersonal relationships (e.g. Finkel & Campbell 2001, Tangney et al. 2004). Succinctly summarizing these findings, Schmeichel (2005) asserts that self-control facilitates success in life, yet self-control often fails (e.g. Baumeister et al. 1994).

Egregious failures of self-control are often observed in circumstances including addiction, crime, or mental and physical illness (Schmeichel 2005; Baumeister et al. 1994; Davidson, et al. 2000; Gottfredson & Hirschi, 1990). Less profound self-control failures often result from (1) divided attention or from (2) previous acts of self-control. Because good self-control typically requires focused attention and awareness, distracted or divided attention undermines, and thus further depletes, self-control. For example, dieters have been shown to be more likely to inhibit food intake when they focus attention on dietary goals, while distracting attention away from dietary goals disinhibits food consumption (e.g. Ward & Mann 2000). Self-control failure also results from previous acts of self-control. That is, exercising self-control appears to deplete a limited internal resource, such that subsequent acts of self-control suffer as a result (e.g. Baumeister et al. 1998; Muraven & Baumeister 2000). Returning to the dieter example, it has been robustly demonstrated that dieters eat more immediately following the exertion of self-control in another domain (e.g. Kahan et al. 2003; Vohs & Heatherton 2000). Relating this to KW, it seems

plausible that complying with a multi-dimensional contract rewarding both creativity and quantity requires a greater degree of self-control than complying with a contract rewarding only quantity.

Recall that, while investigating the decision-influencing and decision-facilitating roles of accounting contracts as strategic performance management systems, KW address whether offering different incentives can allow prospective employees to signal their type, and thus allow employers to sort employees by type. The success of KW's creativity-incentivizing contract in producing *initial* creative output, followed by diminishing creative output over time indicates participants are depleting an internal resource, which I argue from the above discussion is Working Memory (or a component of WM).⁵ Fortunately theory tells us explicitly that the internal limited energy resource drawn down while engaging WM, CE, Self-Control and other similar cognitive functions is a pooled and shared resource; therefore, WM, CE and self-control depletion are used interchangeably for the remainder of the paper.

Related to creative productivity, since "...cerebellar modulations of cortical activity associated with working memory [are hypothesized to be] the fundamental sources of creative and innovative solutions" (Vandervort, Schimpf & Liu 2007, p. 10), *it seems reasonable that reductions in WM resulting from the exercise of self-control may impact creative ability*. That is, exercising self-control in order to focus on producing only high-creativity works depletes the WM available for creative tasks. The above discussion implies a decrease in WM will result in lower creative production as WM becomes depleted. This leads to the underlying research question, which asks whether WM depletion causes participants compensated on creativity-weighted productivity to produce less than participants compensated on quantity alone. This leads to the following formally stated RQ and replication hypothesis:⁶

⁵ One plausible alternative explanation for the contracts' sorting benefits is that participants with many initial ideas self-selected into the creativity-incentive contract, produced their ideas, and depleted their idea stores. Given the relatively uniform decrease around the same point (between 8 and 10 creative ideas) in the experiment, for this explanation to hold, people sorting into creative contracts would have to have known ex ante they had eight or so high-creativity ideas available. They would then have to manage their ideas during the ¼ of their total experimental time spent in a pre-session "practicing". Finally, they would have had to produce their ideas without coming up with new ones.

⁶ In the interest of full disclosure there were initially two replication hypotheses. Unfortunately, due to a clerical error, the order in which experimental participants submitted their puzzles was lost when we randomized the puzzles for the raters. I am therefore unable to test the following hypothesis, stated here for the sake of thoroughness. *Untestable hypothesis: "Initial creativity will be greater under the creativity-weighted contract than under the quantity-only contract."*

RQ: Is the difference KW find between the cumulative creativity-weighted productivity condition and the quantity-only productivity condition the result of Working Memory (WM) depletion?

H1 (replication/baseline hypothesis): Absent WM replenishment, cumulative creativity-weighted output under the quantity contract will be greater than cumulative creativity-weighted output under the creativity-weighted contract

As discussed, acts of self-control consume working memory (Baumeister et al. 1998; Muraven & Baumeister 2000), and multiple studies have shown that WM is a depletable resource (e.g. Muraven, Tice & Baumeister 1998; Vohs & Schmeichel 2003; Vohjs, Baumeister & Ciarocco 2005). In particular, WM is depleted when participants maintain compliance with a high-cognitive effort tasks. An illustrative example of a “high cognitive effort task” from Gilbert, Krull & Pelham (1988) is summarized as follows:

Participants are asked to view a movie with no sound for 6 minutes. In one condition, participant A is simply asked to watch the film. In the second condition, participant B is asked to watch the film and ignore the random words that are projected into the bottom of the screen (these words appear in both conditions, but the first participant is not told to “do” anything about them). After six minutes of performing this task, on average, participant B performs remarkably worse on tasks requiring self-control.

In this study I argue that an example of a high cognitive effort task is participant compliance with the creativity-weighted productivity contract of KW (2010). This is reasonable because creative output is the highest order function of the human mind (De Dreu et al. 2012); In fact, it is arguably our creative abilities which make us human. Yet WM depletion resulting from high cognitive effort tasks, as with WM depletion resulting from acts of self-control, has been shown to be replenishable in the laboratory (e.g. Winters 2010; Gailliot et al. 2007). Recalling that sufficient free WM is necessary for creative improvisation (e.g. De Dreu et al. 2012), this leads to this study’s main hypothesis.

H2: The effects of WM replenishment on creative output are more positive under the creativity-weighted contract than under the quantity-only contract.

III. METHOD

This section first discusses the operationalization of the WM construct, and then discusses the experimental design and randomized treatment. The method of determining creativity ratings and differences in my experimental design and that of KW conclude this section. The following section presents results.

Operationalization of Cognitive Resource Replenishment (specifically WM)

The human brain is a particularly expensive piece of anatomical equipment – especially when one takes an evolutionary perspective. For instance, the (human) brain takes up only 2% of the body’s mass, but uses well over 20% of the body’s energy; most of the energy consumed by the brain is absorbed in the form of glucose (Dunbar, 1998). Time, exercise and glucose administration all replenish WM (Gailliot et al. 2007). This study follows Gailliot et al. (2007) in operationalizing cognitive replenishment of WM using glucose administration. Specifically, glucose is administered to treatment participants in the form of a sugary beverage, and the placebo is administered in the form of a Splenda™ sweetened beverage. More specifically, in order to address H2, I modify the experimental procedures of KW as follows. As discussed earlier, I employ a 2x2 between-subjects experimental design allowing self-selection into contract type and manipulating WM replenishment. In line with KW, I operationalize contract type in basing performance-contingent compensation on creativity-weighted productivity or quantity only. In exploring the effects of WM depletion on the KW creativity-weighted productivity results, I manipulated whether participants receive glucose or a placebo prior to the “start” of the experiment.

Following Gailliot et al. (2007), participants were instructed to fast (i.e. refrain from eating or drinking anything other than water, unsweetened coffee or tea) for 3 hours before they arrive at the experiment. The reason for this is that eating causes idiosyncratic fluctuations in glucose levels to occur throughout the day. Refraining from eating allows glucose levels to stabilize by the time of the experiment, reducing exogenous glucose blood level variations during the experimental treatment. In addition, glucose is absorbed into the bloodstream at a rate of about 30 calories per minute, and after 10 minutes it can be metabolized by the brain (Donohoe & Benton 1999). Because research indicates the

time required to significantly reduce WM is approximately 6 minutes of performing a self-control task, it was necessary to administer the glucose prior to participants making their contract decisions. In short, I operationalized and implemented cognitive replenishment by administering glucose or a placebo using the approach of Gailliot et al. (2007; experiment 7, page 330). As discussed, the present study's experiment is based heavily on the work of KW; therefore, although this section stands alone, I occasionally reference the published discussion of their design (see KW, beginning on p. 1673) for nonessential details rather than repeating their methodology section in full here. That is, in general I follow KW's approach in replicating KRW and note differences where the present study differs from theirs in order to allow treatment with glucose rather than defining treatment as the presence/absence of contract choice.

Similar to KRW and KW, I recruited BBAs from a top business school (one MBA). In each of three sessions, 10-15 volunteers participated in an hour-long experiment, yielding a total of 38 participants (27 female) from a wide variety of undergraduate business majors. I took particular care to copy the instructions from KW precisely to the point of using their instrument, which I both requested and received from one of the authors (see acknowledgements). One important point is that I used the instructions which are common to both studies, in particular stating, "While we do not place any rules on the kinds of rebus puzzles you can submit, we value both the number of different puzzles you construct (i.e., quantity) and the creativity of those puzzles (i.e., puzzles that are original ideas, innovative, and clever)." As was the case with KRW and KW, my interest is in the incremental effects of performance-based compensation rather than the goal-setting properties of demand effects (Bonner & Sprinkle 1991). The remainder of this section presents the methodology, procedures used and differences in experimental designs between the current study and KW's.

In a 2x2 between-subjects design, I measure participants' choice of contract as that should provide an indication of type, and I am interested in the selection benefits that can occur with sorting contracts. I manipulate whether the participant receives cognitive replenishment operationalized using glucose or a placebo operationalized using SplendaTM. As with KW, participants worked separately and the puzzle

submission process was designed to maintain the order of the cards.⁷ I also followed KW's pay scheme in anchoring the low (\$5) and high (\$45) endpoints, thus allowing me to vary the pay basis but still hold the expected magnitude constant. As stated by KW (2010; p. 1675, footnote 7),

Similar to KRW, we achieve this equivalence by basing the low and high payments on the lowest and highest performers within each contract. However, these contracts are not "tournament" schemes as defined by Bonner et al. (2000), because with the sole exception of the highest-scoring participant, every other participant can earn greater compensation by increasing the compensated performance measure, irrespective of ordinal ranking.

I also followed KW's procedures with one primary difference. In order to make sure the glucose had time to reach the brain prior to the end of the 20 minute period, I administered the sweetened beverage (lemonade Kool-Aid made with Splenda or with Sugar) while participants were reading instructions, but *after* they completed the pre-experimental questionnaire. There are two reasons for this temporal difference: (1) it is critical to be aware of any medical issues that might cause a participant to have a poor medical outcome in reaction to being given a sugary beverage (i.e. history of diabetes, phenylketonuria, etc.), and (2) in order to assess the impact of glucose on participants' responses, it was necessary to have them complete the task while the glucose was taking effect, then complete the post-experimental questionnaire. In this study, participants answered the questionnaire in line with KRW (i.e. post-experiment) rather in line with KW (i.e. following the practice period).⁸ There were no negative medical reactions, although one diabetic required me to break strict randomness since that person had to receive the placebo-sweetened rather than the glucose-sweetened beverage.⁹ Once all the puzzles were collected, it was necessary to determine average creativity ratings.

In order to determine creativity ratings, two undergraduate student volunteers and one administrative staff (n=3) independently rated the creativity of the 967 rebus puzzles produced by the experimental participants during the compensated phase of the experiment. Raters were blind to treatment condition

⁷ As was mentioned earlier, a clerical error resulted in my losing order in my cards. Although this does not mean I am unable to test my theory, it does mean it is challenging to understand what might have occurred should I be unable to replicate KW.

⁸ Glucose takes 10 minutes following oral consumption to become available to the brain (bioavailability) for use as an energy source (Donohoe & Benton 1999). Therefore, investigating whether the effect of cognitive resource replenishment is greater under the creativity-weighted contract than under the quantity-only contract requires post-experimental administration of the questionnaire. Hence, I follow KRW in administering the post-experimental questionnaire (PEQ) following the experiment. As discussed in KW, who successfully replicates KRW's results, the timing of questionnaire administration seems unlikely to interfere with inferential results.

⁹ Inferential results remain consistent regardless of whether this individual is present in the analyses.

and participant identity, and each rater was paid \$50 for their time. The size of my experiment allowed each rater to rate all puzzles in the same order. In addition to the number of raters (KW used 40) there are a few additional differences in experimental design that bear mentioning.

Three differences in experimental method between my study and KW's have now been mentioned; any of these may cause inferential challenges. First all my participants were asked to fast for at least three hours prior to commencing the experiment, and all of them responded that they had indeed fasted at least three hours, to reduce (or eliminate) idiopathic blood sugar changes (Gaillet et al. 2007).¹⁰ Second, I have only three raters, and the interrater reliability scores range from a low of -0.0702 to a high of -0.0339 where interrater reliability scores are calculated as the Kappa statistics between each rater's ratings and (1) each of the other two rater's ratings, then (2) the average of the other two rater's ratings. Although time and cost considerations kept me from administering precisely the same rating scheme as KW, it certainly seems desirable given a larger and better funded study to access the remedy they implemented in dropping the lowest correlated rater within each session.

The third difference in the methods section bears mentioning at this point. KW report the range of the average ratings for their entire distribution of averages; it is 1.44 to 8.89. For this study, the range for the entire distribution of averages is 6.22 to 7.77. This "bunching" of the creativity ratings could mean the raters were unusually generous, that the puzzles were all of uniform type, or that the puzzles were simply that much better than those produced in KW's experiment, among other reasons.¹¹ In other words, there may be significant differences in the rating populations, the participant populations, or both.

IV. RESULTS

This section tests first for successful replication of Kachelmeier and Williamson (2010) then tests this study's hypotheses. Recall the research question investigates why KW participants' cumulative creative output differs between the creativity-weighted and quantity-only contracts. Specifically, the research

¹⁰ Differences in glucose levels can lead to significantly different outcomes even across short time windows.

¹¹ Spot analysis indicates this last (i.e. more creative overall puzzles) is a highly unlikely reason for the "bunching."

question asks whether multidimensional contracts rewarding participants for creativity-weighted output deplete pooled cognitive resources to the detriment of measurable creative output.

Contract Selection Benefits

Prior to testing their hypotheses, KW explore three necessary conditions for assessing the benefits of contract self-selection; these are that: (1) participants perceive differences in their ability to produce high-creativity rebus puzzles, (2) meaningful variation in actual creative output exists, and (3) self-assessed creative potential and output are positively correlated. Table 1 (Panel A) provides univariate results and descriptive statistics. A copy of KW's Table 1 (see Table 1, Panel B of this study) is provided for convenient comparison. Similarly, Table 1 (Panel C) presents p-values comparing KW's means to those of the current study.

Following KW, I elicit participants' self-assessed ability to produce highly creative rebus puzzles using a nine-point Likert scale questionnaire from KW, which I administer *at the end* of the experiment. I find support for KW's finding that people who believe themselves more creative are significantly more likely to choose to enter into a contract rewarding creativity. Specifically, approximately 60% of participants (23 of 38) choose the creativity-weighted contract. Comparison of means reveals participants selecting the creativity-weighted contract believe themselves more creative than participants selecting quantity-only contract ($t = 2.25$; two-tailed $p < .031$). Moreover, participants rating their creativity above the median (median = 5) disproportionately choose the creativity-weighted contract (7 of 8), while participants rating their creativity at the median ($n = 16$) similarly choose the creativity-weighted contract, although at a somewhat lower rate (11 of 16) when compared to participants above the median.¹²

Untabulated logistic regression results support KW's findings that participants become more likely to choose the creativity-weighted contract as self-perceived creativity increases. In sum, these analyses reproduce KW's in that results suggest participants selecting the creativity-weighted contract perceive themselves as having above-average creativity, while participants selecting the quantity-only contract

¹² I follow KW by regressing a logit of participants' propensity to choose the creativity-weighted contract on self-assessed creativity potential (Coefficient = .593; $p = .0215$, one-tailed).

perceive themselves as having below-average creativity. That is, the contracts successfully sort participants based on their subjective beliefs regarding their creative potential.

Next, I assess whether participants reveal variation in creative output. I define high-creativity output as the number of high-creativity puzzles produced by each participant, where high-creativity puzzles are those ranking in the top quartile (here > 7.33). Figure 1 demonstrates the relative proportion of high-creativity puzzles per condition, while Figure 2 provides a distribution of high-creativity output by participant. As Figure 2 demonstrates, variation in high-creativity production is well-dispersed.

<<< INSERT FIGURE 1 ABOUT HERE >>>

<<< INSERT FIGURE 2 ABOUT HERE >>>

Although these results are generally in line with KW's findings, a couple of interesting observations bear mentioning. First, the number of overall puzzles produced in the quantity-only condition vs. in the creativity-weighted condition does not appear to be as great as that of KW. Second, the difference in the number of highly creative puzzles produced by condition appears greater than in KW. These observations are discussed in greater detail and tested in the tests of hypotheses below.

Next, I follow KW by constructing two "pseudo-samples" where I divide participants at the median into those producing a high number of high-creativity puzzles (high-creativity group) and those producing a low number of high-creativity puzzles (low-creativity group), where the median number of high-creativity puzzles is 4. I remove the seven participants who produced the median number of high-creativity puzzles ($n=4$) from this set of analyses. I find that high-creativity group participants also perform better in terms of mean creativity ratings when compared to low-creativity group participants (high-creativity group mean creativity rating = 7.20 vs. low-creativity = 6.78; $p < 0.004$, two-tailed), but they do not produce a greater overall number of puzzles in aggregate (high-creativity $N = 28$ vs. low-creativity $N = 24.5$; $p > 0.35$, two-tailed). Yet, as in KW, no significant statistical correlation exists between participants' self-perception of their potential creativity and the aggregate number of high-creativity puzzles each participant produces ($r = 0.12$; $p > 0.47$, two-tailed). These results indicate participants may find it difficult to accurately predict and match their aggregate production ability with

their contract selection preference. Unlike KW, this analysis indicates that I may lack the statistical power to discern creativity-based advantages in aggregate high-creativity production because participants may not have had the foresight to select contracts matching their ability to produce with their actual differences in creative production (see KW, p. 1680, footnote 15).

Hypothesis Tests

Hypothesis Tests: H1

Recall that H1 is the replication hypothesis; formally stated as, “Absent WM replenishment, cumulative creativity-weighted output under the quantity contract will be greater than cumulative creativity-weighted output under the creativity-weighted contract.” Surprisingly, I do not find support for H1 in that the experimental results show aggregate creativity-weighted productivity is greater under the creativity-weighted contract than under the quantity contract.¹³ Specifically, I find that participants in the creativity-weighted contract produce higher rated puzzles (7.27 vs. 6.77; $p < 0.007$, two-tailed) and more high-creativity puzzles (6.67 vs. 3.00; $p < 0.023$, two-tailed) puzzles.

Upon reflection, there are at least four possible explanations why my experiment failed to replicate results from KW (2010). First, as was noted previously, KW used 40 raters across four sessions. Thus, although KW were able to drop the lowest correlated rater from each session, I was unable to follow this design process as this experiment had only three total raters. As mentioned in the Differences in Experimental Method section above, this led to poor interrater reliability scores.¹⁴ Poor interrater reliability leads to noise in the measure that I am unable to resolve with only three raters; a possible solution is to have additional raters rate the puzzles produced.

Second, as is shown in the following graphic (Figure 3), the effects of familiarity differ between the quantity and creativity-weighted contracts. Familiarity is simply a measure of participants’ self-assessed pre-experimental familiarity with rebus puzzles. As with KW, familiarity was captured using a nine-point Likert scale stating, “I was familiar with rebus puzzles before the experiment.” Regarding the graphic,

¹³ “Surprisingly” because these results are inconsistent with both KRW and KW

¹⁴ Furthering this conjecture is that there appears to be some “bunching” of the ratings near the “high” end of the scale in that the average rating for KW is near 5.3 while the average rating for this experiment is closer to 7.3. This difference is statistically significant (see Table 1).

working within only the Placebo condition (the replication condition), it appears familiarity affects creativity ratings under the *quantity* contract, while familiarity does *not* seem to affect creativity ratings under the *creativity-weighted* contract. Specifically, it appears familiarity matters in the Placebo condition under the quantity-only contract while it does not under the creativity-weighted contract. Particularly interesting are the apparent differences in the lower ranges of observations between conditions. For instance, consider the mean creativity rating for participants with familiarity less than 3. Note that mean creativity rating for these participants in the quantity-only contract is significantly lower than for participants in the creativity-weighted contract (6.46 vs. 7.28; $p < 0.001$). Importantly these participants also produced significantly fewer high-creativity puzzles (1.75 vs. 7; $p < .052$).¹⁵ In sum, these statistics and visual observation of Figure 3 indicate the possibility that creativity ratings increase with familiarity under the quantity-only contract but not under the creativity-weighted contract. Since KW's participants have higher familiarity across virtually all conditions (see Table 1, Panels A-C), it is possible that familiarity alone induced significant noise into my experimental results.

<<< INSERT FIGURE 3 ABOUT HERE >>>

Third, it is possible that having participants fast prior to the experiment may have differentially influenced their ability to produce creative works. There is some marginal support for this conjecture which I discuss further in the supplementary analysis section. Finally, it is possible that the small sample size or other measures that differ between my subject pool and that of KW could have negatively impacted my ability to replicate KW.¹⁶

Hypothesis Tests: H2

Despite the fact that the above results demonstrate this experiment did not replicate the results from KW, the main hypothesis from this paper remains testable. Recall that H2 is formally stated as, “The effects of WM replenishment on creative output are more positive under the creativity-weighted contract

¹⁵ Note these analyses are somewhat speculative since quantity-only $n=4$ and creativity-weighted $n=6$ for these conditions. Ten bootstraps of this (5000 replications each w/replacement) confirms this is speculative in that p-values become uniformly non-significant and range from 0.172 to 0.540.

¹⁶ These include possible differences in participant characteristics such as participant IQ or Foreign/Domestic composition of the subject pool. I leave these considerations for future analysis.

than under the quantity-only contract.” I therefore proceed to test H2 by following KW in using an analysis of covariance (ANCOVA) to investigate the main and interactive effects of contract type (i.e. quantity-only vs. creativity-weighted) and cognitive resource replenishment (i.e. glucose vs. placebo) on the DVs: (1) Aggregate Creativity-Weighted Productivity (Table 2, Panels A to C) and (2) High-Creativity output (Table 2, Panels D to E) defined as above to be the number of puzzles produced in the top quartile of all puzzles produced (here > 7.33).¹⁷

In both cases, the covariate used is the measure of participants’ self-assessed familiarity with rebus puzzles. As with KW, familiarity was captured with a nine-point Likert scale stating, “I was familiar with rebus puzzles before the experiment.” As with KW, I find familiarity statistically significant in all analyses. Removing the variation associated with familiarity helps me isolate the impact of creativity from pre-experimental exposure to rebus puzzles.¹⁸

Recall that I theorized the effects of Working Memory (WM) replenishment, as operationalized using glucose, are greater for the creativity-weighted contract than for the quantity-only contract. If true, this would result in the following pattern whether Aggregate Creativity-Weighted Productivity or High-Creativity output is on the y-axis:

<<< INSERT FIGURE 4 ABOUT HERE >>>

As is shown in the following graphs and tables, actual outcomes do not support this prediction. In fact, the only significant effect from these ANCOVAs with aggregate creativity-weighted productivity (Panel A) and total high-creativity output (Panel D) as the dependent variables is the main effect of contract type.

<<< INSERT TABLE 2 ABOUT HERE >>>

As with the replication hypothesis discussed above, creativity-weighted participants significantly outperform their counterparts in the quantity-only contract in both Panel B ($F 4.68$; $p < 0.04$, two-tailed)

¹⁷ KW also seek to explain differences in total puzzle output. In no case is the difference in total puzzles produced significant at traditional levels (lowest $p > 0.38$, two-tailed).

¹⁸ That said, following KW, a separate ANOVA with familiarity (untabulated) as the DV shows no main effect or interaction of Contract Type or Glucose, yet Glucose is “close” ($F 2.76$; $p 0.1061$). Because inferential results remain unchanged whether using an ANOVA or if familiarity is used in an ANCOVA, I follow KW in these analyses.

and in Panel E ($F 5.78$; $p < 0.023$, two-tailed). Although neither glucose nor the interaction between glucose and contract type are significant, the pattern of actual results Table 1 (Panel A and Panel D) surprisingly demonstrates the reverse of my predictions. I discuss this potentially interesting occurrence in the supplementary analyses next.¹⁹

Hypothesis Tests: Supplementary Analyses

Without venturing too far afield, some explanation for the pattern demonstrating a possible reversal of the results from KW is in order. The Post Experimental Questionnaire (PEQ) provides a possible explanation in that two PEQ questions show significant differences between the Glucose participants and the Placebo participants. These are Anxiety (Glucose 3.21 vs. Placebo 4.47; $p < 0.031$, two-tailed), Time Pressure (Glucose 3.63 vs. Placebo 4.89; $p < 0.057$, two-tailed). I follow these up with two separate ANOVAs (untabulated) investigating the main and interactive effects of glucose and contract type on the DVs Anxiety and Time Pressure. Neither contract type nor the interaction of contract type and glucose is significant in either case, but glucose has a significant main effect on Anxiety ($F 4.09$, $p < .052$, two-tailed) and Time Pressure ($F 3.21$; $p < .082$, two-tailed).

I followed up these results by reviewing a number of medical journal articles that indicate glucose does indeed lower feelings of time pressure (i.e. tension) and reduce anxiety (e.g. Thayer 1987; Benton & Owens 1993) at least in the short term. These studies find that sugar is associated with decreased tiredness and tension *for the first hour* following administration, but leads to increased tiredness and reduced energy following the first hour. I therefore hypothesize that, although glucose administration may have replenished Working Memory, it had a countervailing (and overriding) effect on participants' motivation to produce under the creativity-weighted contract.

To demonstrate how this may have affected my experimental results, consider a situation where all participants had been given glucose within an hour of presenting for the experiment. Next, consider an extension of the lines associated with the output of these participants from Table 2, Panel A. Assuming

¹⁹ Assuming these predictions would become significant given a large enough sample size (*please note I would not attempt such further analyses in a paper prepared for journal submission, but was advised this would be helpful for this summer paper*), I provide results in Table 2 (Panels B, C, E and F) that analyze the interactions "as if" they were significant, albeit in the opposite direction from those in KW.

creative production was influenced by glucose over time, that graphic may have appeared as follows. In this graphic, the x-axis represents time, the vertical dashed line represents the hypothetical end of the present experiment, the central portion represents a non-fasting participants' potential state depending on the last time they ate, and the solid vertical line represents the start of the KW experiment assuming participants had consumed some form of glucose within the last 1-2 hours.

<<< INSERT FIGURE 5 ABOUT HERE >>>

Although this supplementary analysis leads to a somewhat speculative proposition, it provides an additional avenue for future research and indicates that at least four additional considerations need to be taken into account for research of this type. First, when attempting to replenish working memory, it is critical for scientists to test that WM was actually restored. This can be accomplished using one of the multiple tests of WM capacity from the Psychology literature (e.g. Gailliot et al 2007). Second, in experiments involving measuring creative output using rebus puzzles, maintaining order in the puzzles produced allows the experimenter a much better understanding of whether unforeseen interactions affect participants' output over time. Third, different methods of replenishing working memory such as exercise may be more appropriate than glucose. Fourth, when choosing to administer glucose, it may benefit the experimenter to take blood glucose levels from participants in order to measure the level of glucose both pre and post experiment.

V. CONCLUSIONS

Research into incentivizing creativity is valuable to accounting researchers because such research relates to both the decision-facilitating and decision-influencing roles of managerial accounting. Creativity-incentivizing contracts act to facilitate employee decisions because incorporating pay-for-creative output into employee contracts directs employee attention toward increased creative output. Results from this experiment illustrate the importance of considering physiological changes when making inferences based on experimental laboratory results. This may be particularly true in cases where participants are constrained by the experimental environment from maintaining what should be *ceteris paribus* conditions with the "real world."

The present research contributes to the multiple prior studies that have investigated and demonstrated the benefits of using incentive contracts for selection purposes (e.g. Chow 1983; Waller & Chow 1985; Cadsby et al. 2007). In line with Kachelmeier et al. (2008; KRW) and Kachelmeier and Williamson (2010; KW), this study helps extend our understanding of experiments rewarding on a single performance dimension to those rewarding on multiple performance dimensions by examining the effects of contract selection on creativity-weighted production in a multi-dimensional setting. In order to extend KRW and KW, I experimentally investigate contract selection in a multidimensional setting. Specifically I examine the effects on creativity-weighted productivity of allowing participant self-selection into either a contract rewarding creativity-weighted productivity or a contract rewarding quantity-only productivity.

In contrast to KW, I do not find that participants' self-perceptions lack accuracy over the long-run; indeed, contrary to KW, I find that participants who self-select into the creativity-weighted contract produce greater overall creativity-weighted output as measured by (1) their average creativity-weighted output, and (2) the number of highly creative works produced per participant. On the other hand, this study successfully replicates KW's finding that participants choosing the creativity-weighted contract perceive themselves to have greater creative potential, thus providing further support for the conjecture that there is value in using incentive contracts as sorting devices. This is consistent with KW's premise that creativity-weighted contracts are able to attract more creative workers. Differences in our experimental designs may explain why these results are not congruent. I discuss these next.

This experiment applies theory from psychology and physiology to support the hypothesis that focusing on producing creative output results in the depletion of Working Memory. Depletion of Working Memory (WM) is important because WM is the ability to direct and control attention (Baddeley & Hitch 1974), and WM-directed attention is a requirement for creative performance (De Dreu et al. 2012). I replenished WM using glucose administered to participants after they had fasted (i.e. did not eat or drink) for three hours pre-experiment (e.g. Gailliot et al. 2007).

The results were surprising. Not only was I unable to replicate results from KRW and KW, but I also found almost precisely the opposite results from those predicted in this study's main hypothesis. That is,

glucose administration appears to *decrease* participants' creative output under the creativity-weighted contract, while *increasing* participants' output under the quantity-only contract. Upon further reflection, these results may be explained by the short-term effects of glucose on the tension and incentives of participants who are in a fasting state. Specifically, it is possible these manipulations induced a "sugar high" for participants. Additional analyses confirm that participants in the glucose conditions were less likely to feel anxiety and time pressure under the creativity-weighted contract, but not under the quantity-only condition. Thus it is conceivable that glucose administration negatively impacts creative output under a multidimensional contract incentivizing creativity while either having no effect or a positive effect on creative output under a quantity-only contract.

The research question, however, was not related to glucose administration per se, but to the idea that physiological mechanisms must be taken into account when designing and running experiments that compare multi-dimensional and single-dimensional contracts. In this regard the present study responds to calls by KRW and KW requesting an expansion of the experimental accounting literature on performance-based incentives beyond the traditional interest in quantity alone. Respecting that call, I believe that - although the experimental outcomes were opposite those predicted - the results clearly demonstrate that considerations of human psychology *and physiology* are likely to be important as accounting academics seek to improve our understanding of multi-dimensional incentive contracts using laboratory methods.

VIII. REFERENCES

- Amabile, T. M. 1996. *Creativity in Context: Update to the Social Psychology of Creativity*. Boulder, CO: Westview Press.
- Baddeley, A.D. (2007). *Working Memory, Thought, and Action*, Oxford University Press.
- Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, 74, 1252-1265.
- Benton, David, and Deborah Owens. "Is raised blood glucose associated with the relief of tension?." *Journal of psychosomatic research* 37, no. 7 (1993): 723-735.
- Chang, Linda, and Bill Birkett. "Managing intellectual capital in a professional service firm: exploring the creativity–productivity paradox." *Management Accounting Research* 15, no. 1 (2004): 7-31.
- Chow, C. W. 1983. The effects of job standard tightness and compensation scheme on performance: An exploration of linkages. *The Accounting Review* 58 October: 667–685.
- Chen, Clara Xiaoling, Michael G. Williamson, and Flora H. Zhou. "Reward system design and group creativity: An experimental investigation." *The Accounting Review* (2012).
- Christ, Margaret, Scott Emett, William Tayler, and David Wood. "To Compensate or Control? Motivating Employees in a Multidimensional Task." *Motivating Employees in a Multidimensional Task* (July 27, 2011) (2011).
- De Dreu, Carsten K. W., Bernard A. Nijstad, Matthijs Baas, Inge Wolsink, and Marieke Roskes. "Working Memory Benefits Creative Insight, Musical Improvisation, and Original Ideation Through Maintained Task-Focused Attention" *Personality and Social Psychology Bulletin* (2012) 38: 656
- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125, 627–668.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.
- Deci, E. L., & Ryan, R. M. (1987). The support of autonomy and the control of behavior. *Journal of Personality and Social Psychology*, 53, 1024–1037.
- Donohoe, R. T., & Benton, D. (1999). Blood glucose control and aggressiveness in females. *Personality and individual differences*, 26(5), 905-911.
- Eisenberger, R., Pierce, W. D., & Cameron, J. (1999). Effects of reward on intrinsic motivation—Negative, neutral, and positive. *Psychological Bulletin*, 125, 677–691.

Eisenberger, R., & Rhoades, L. (2001). Incremental effects of reward on creativity. *Journal of Personality and Social Psychology*, 81, 728–741.

Eisenberger, R., Rhoades, L., & Cameron, J. (1999). Does pay for performance increase or decrease perceived self-determination and intrinsic motivation? *Journal of Personality and Social Psychology*, 77, 1026–1040.

Eisenberger, R., & Shanock, L. (2003). Rewards, Intrinsic Motivation, and Creativity: A Case Study of Conceptual and Methodological Isolation. *Creativity Research Journal*, 15, 121-130.

Fallon, Pat, and Fred Senn. *Juicing the Orange: How to turn creativity into a powerful business advantage*. Harvard Business Press, 2006.

Gailliot, M. T. (2008). Unlocking the energy dynamics of executive functioning: Linking executive functioning to brain glycogen. *Perspectives on Psychological Science*, 3(4), 245-263.

Gailliot, M. T., Baumeister, R. F., DeWall, C. N., Plant, E. A., Brewer, L. E., & Schmeichel, B. J. (2007). Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor. *Journal of Personality and Social Psychology*, 92, 325-336.

Hofmann, W., Schmeichel, B. J. and Baddeley, A. D. 2012. Executive functions and self-regulation. *Trends in Cognitive Sciences* 16 (3): 174-180.

Ittner, Christopher D., David F. Larcker, and Madhav V. Rajan. "The choice of performance measures in annual bonus contracts." *Accounting Review* (1997): 231-255.

Jonides, J. and Smith, E.E. (1997). The architecture of working memory. *Cognitive neuroscience* (Rugg, M.D., ed.), pp. 243–276, MIT Press

Kachelmeier, Steven J., and Michael G. Williamson. "Attracting creativity: The initial and aggregate effects of contract selection on creativity-weighted productivity." *The Accounting Review* 85, no. 5 (2010): 1669-1691.

Kachelmeier, Steven J., Bernhard E. Reichert, and Michael G. Williamson. "Measuring and motivating quantity, creativity, or both." *Journal of Accounting Research* 46, no. 2 (2008): 341-373.

Kaplan, Robert S., and David P. Norton. "The balanced scorecard—measures that drive performance." *Harvard business review* 70, no. 1 (1992): 71-79.

Kaplan, Robert S., and David P. Norton. "Using the balanced scorecard as a strategic management system." *Harvard business review* 74, no. 1 (1996): 75-85.

Miller, E. K., & Wallis, J. D. (2009). Executive function and higher-order cognition: Definitions and neural substrates. In *Encyclopedia of Neuroscience* (Vol. 4, pp. 99-104). Oxford, UK: Academic Press.

Miyake, Akira, Naomi P. Friedman, Michael J. Emerson, Alexander H. Witzki, Amy Howerter, and Tor D. Wager. "The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis." *Cognitive psychology* 41, no. 1 (2000): 49-100.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press.

Monsell, S. (2003) Task switching. *Trends in Cognitive Science* 7, 134–140.

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, 126, 247-259.

Ruscio, A. M., & Amabile, T. M. (1999). Effect of instructional style on problem-solving creativity. *Creativity Research Journal*, 12, 251–266.

Simon, H.A., 1961. *Administrative Behaviour: A Study of Decision Making Processes in Administrative Organisation*, 2nd ed. The Macmillan Company, New York.

Simons, R., 1995. *Lever of Control: How Managers Use Innovative Control Systems to Drive Strategic Renewal*. Harvard Business School Press, Boston.

Sprinkle, Geoffrey B. "Discussion of Measuring and motivating quantity, creativity, or both." *Journal of Accounting Research* 46, no. 2 (2008): 375-382.

Stokes, P. D. (2012). Creativity and operant research: Selection and reorganization of responses. In M. A. Runco (Ed.), *Handbook of creativity research*. Cresskill, NJ: Hampton.

Thayer, Robert E. "Energy, tiredness, and tension effects of a sugar snack versus moderate exercise." *Journal of Personality and Social Psychology* 52, no. 1 (1987): 119.

Vandervort, L. (1997). The evolution of Mandler's conceptual primitives (image schemas) as neural mechanisms for space-time simulation structures. *New Ideas in Psychology*, 15, 105–123.

Vandervort, L. (2003). How working memory and cognitive modeling functions of the cerebellum contribute to discoveries in mathematics. *New Ideas in Psychology*, 21, 159–175.

Vandervort, Larry R., Paul H. Schimpf, and Hesheng Liu. "How working memory and the cerebellum collaborate to produce creativity and innovation." *Creativity Research Journal* 19, no. 1 (2007): 1-18.

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

FIGURES

As in KW, I determine whether there is meaningful variation in creative output by producing a histogram of the number of high-creativity puzzles produced by participant. High-creativity puzzles are those in the top quartile of all puzzles produced which, in this study’s distribution, are puzzles with average creativity scores > 7.33 .

FIGURE 1

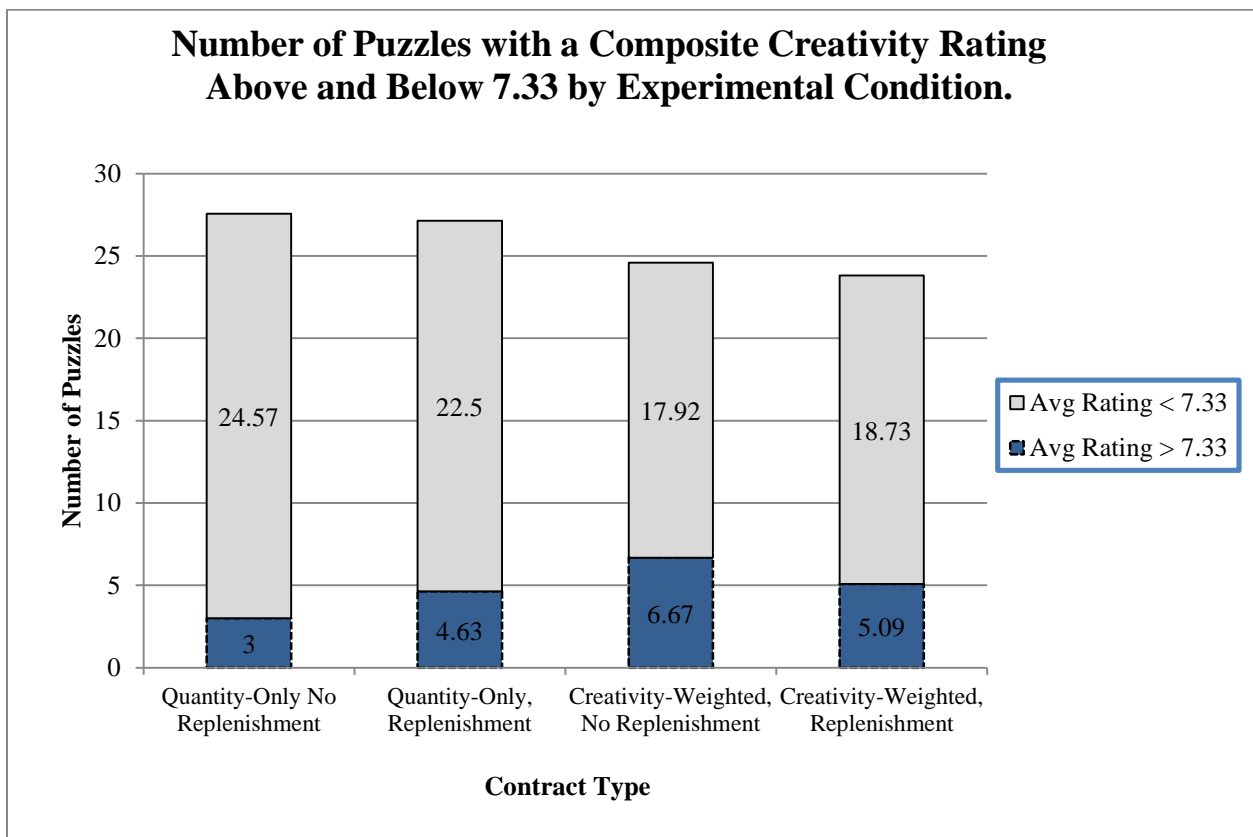


Figure 1 splits the average quantity produced within each experimental condition into high-creativity puzzles and all other puzzles. What is noticeably different in Figure 1 from this study and the comparable figure from KW (Figure 5, p. 1688) is that quantity-only participants in the present study do not appear to produce a much greater quantity of puzzles overall, and they appear to produce fewer high-creativity puzzles. Statistical tests presented in confirm these observations.

FIGURE 2

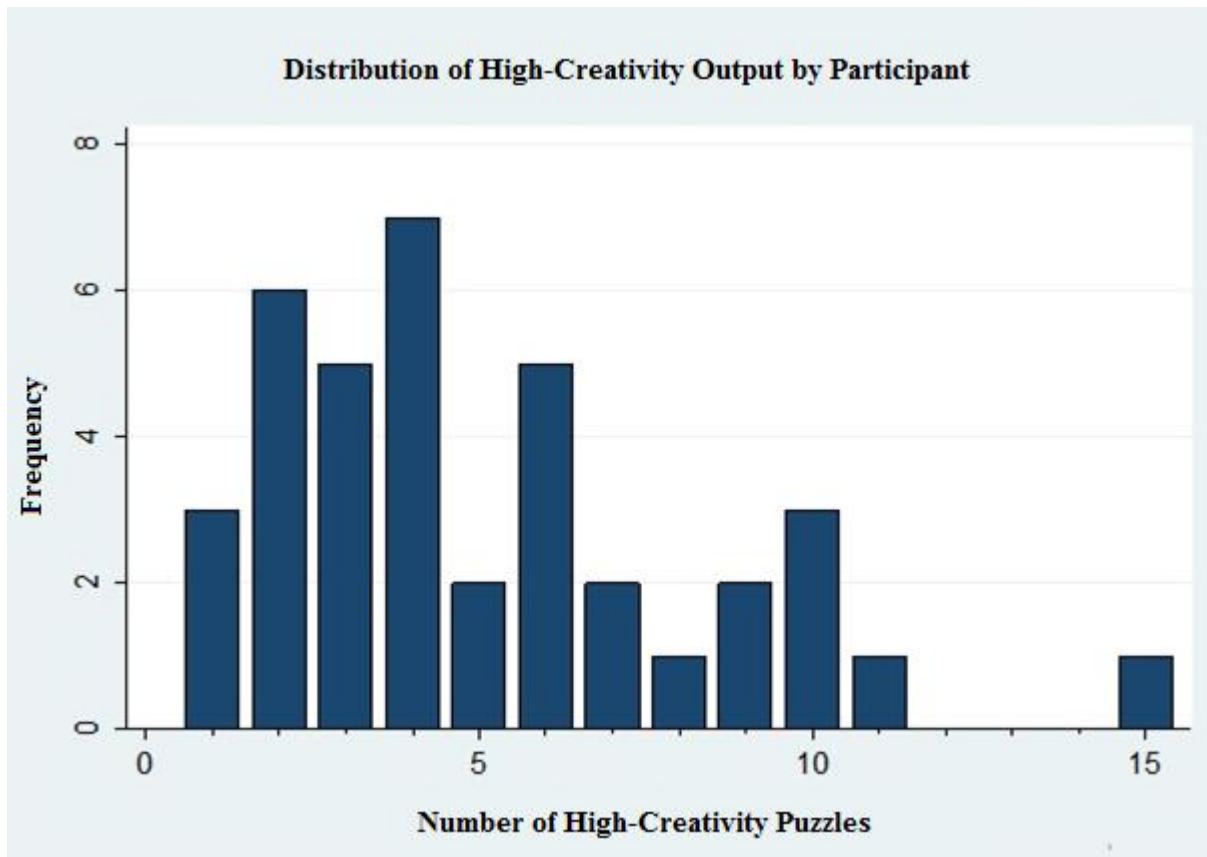


Figure 2 presents a histogram of high-creativity output for participants collapsing across both contracts where the number of high-creativity puzzles produced is on the x-axis, and the number of participants producing those puzzles on the y-axis. As this figure indicates, similar to KW, the variation of high-creativity production is dispersed, with three participants producing two or fewer high-creativity puzzles and one participant producing 15.

FIGURE 3

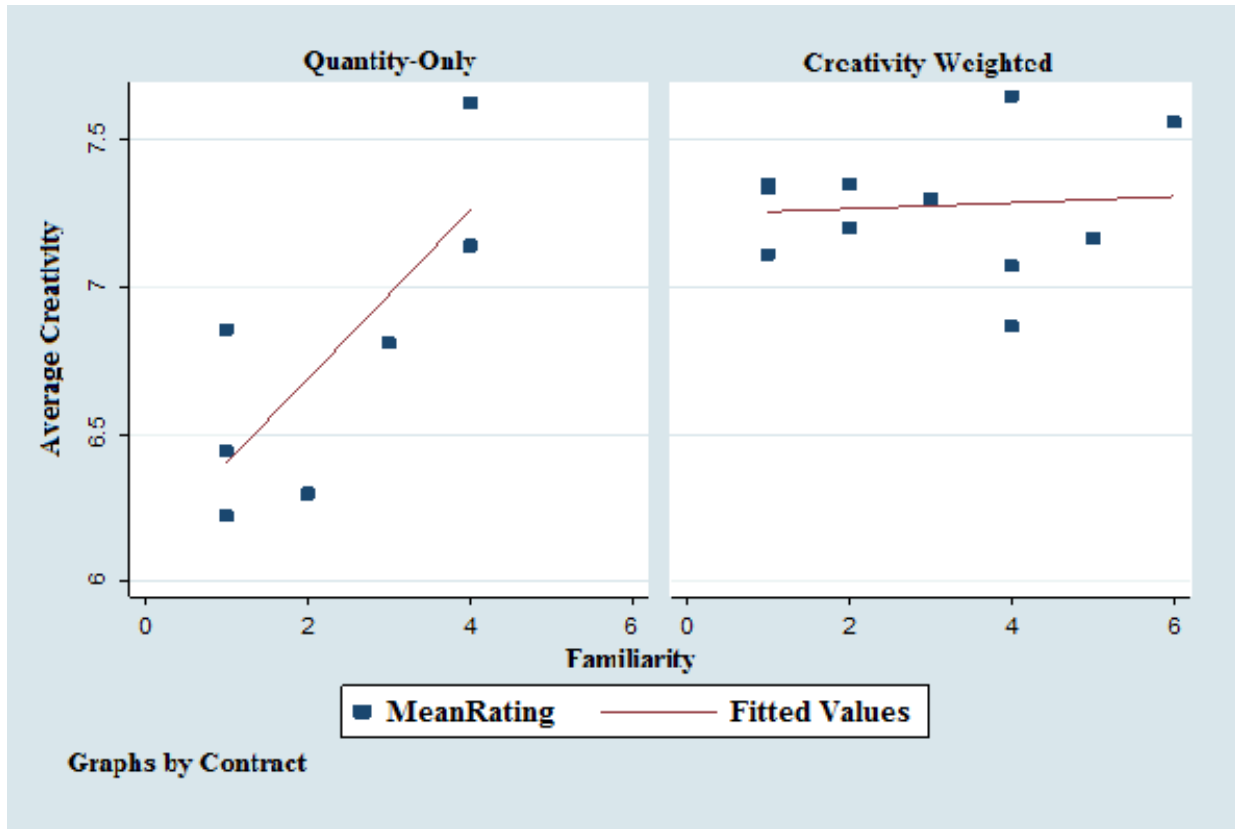


Figure 3 graphs each participant’s average creativity score by contract for participants in the Placebo condition. What seems apparent from these two scatterplots and linear transformations is that the effects of familiarity differ between the quantity and creativity-weighted contracts. Familiarity is simply a measure of participants’ self-assessed pre-experimental familiarity with rebus puzzles. As with KW, familiarity was captured with a nine-point Likert scale stating, “I was familiar with rebus puzzles before the experiment.” Regarding the graphic, working within only the Placebo condition (the replication condition), it appears familiarity affects creativity ratings under the *quantity* contract, while familiarity does *not* seem to affect creativity ratings under the *creativity-weighted* contract. For instance, consider the mean creativity rating between conditions for participants with familiarity less than 3 and note that mean creativity rating for these participants in the quantity-only contract is significantly lower than for participants in the creativity-weighted contract (6.46 vs. 7.28; $p < 0.001$). As noted in Section V, these participants also produced significantly fewer high-creativity puzzles (1.75 vs. 7; $p = .051$).

FIGURE 4

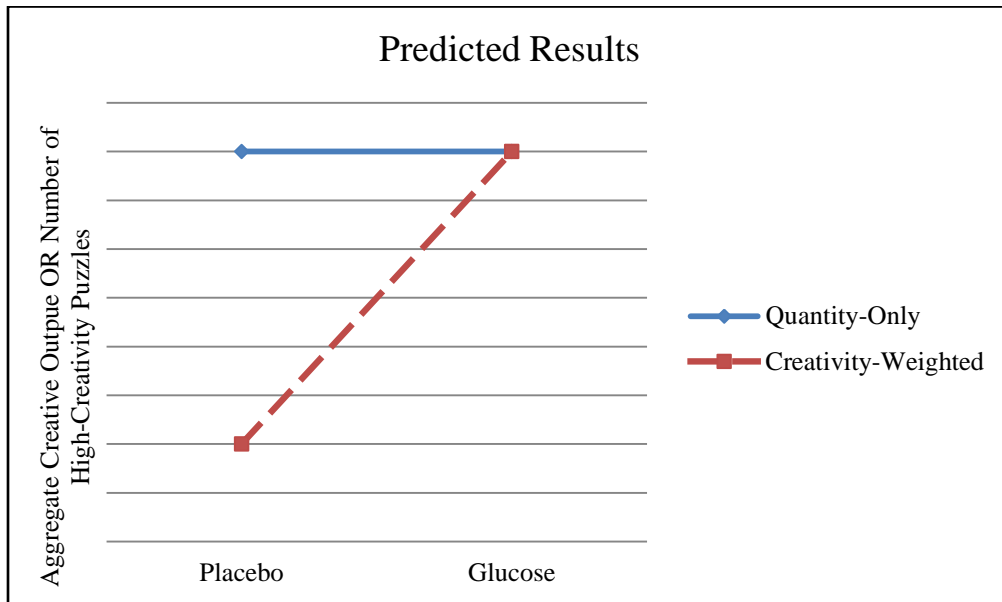


Figure 4 is a graphical depiction of this study's main hypothesis which predicts the effects of WM replenishment on creative output are more positive under the creativity-weighted contract than under the quantity-only contract.

FIGURE 5

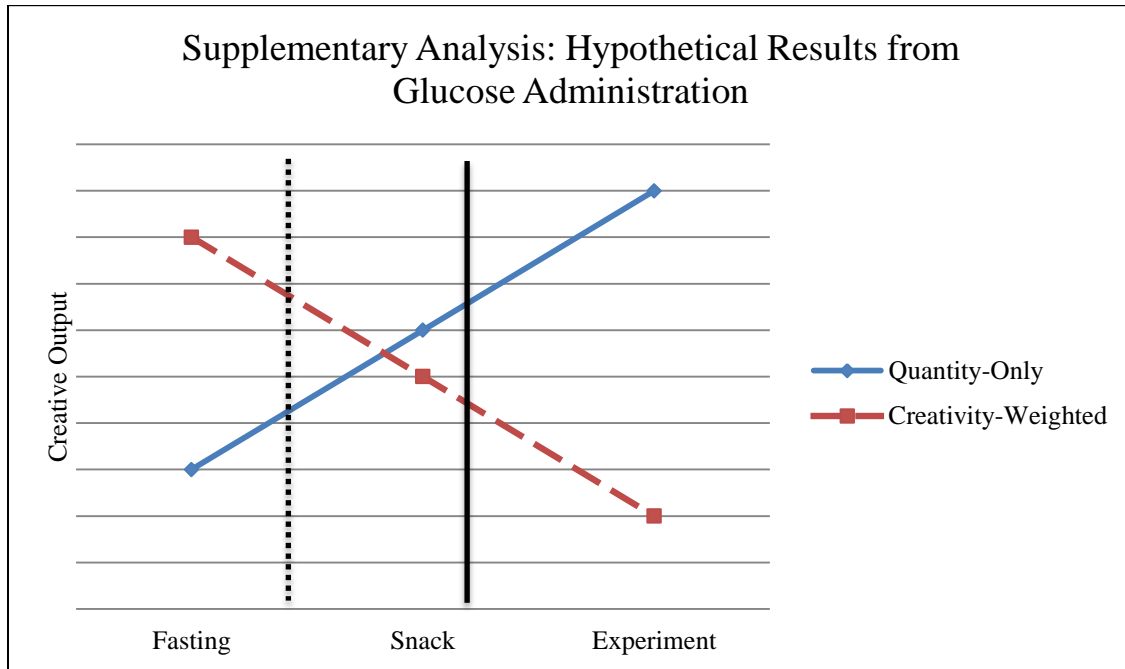


Figure 5 is a hypothetical scenario intended to demonstrate how asking participants to start this experiment in a fasting state may have affected the experimental results. That is, consider a situation where all participants have had glucose (or other simple sugar) in some form within an hour of presenting for the experiment. Next, consider an extension of the lines associated with the output of these participants from Table 2, Panel A. Assuming creative production was influenced by glucose over time, that graphic may have appeared as in Figure 5. Here the x-axis represents time, the vertical dashed line represents the hypothetical end of the present experiment, the central portion represents a non-fasting participants' potential state depending on the last time they ate, and the solid vertical line represents the start of the KW experiment assuming participants had consumed some form of glucose within the last 1-2 hours. Notice that continuing the lines in this hypothetical experiment into a "normal" day (i.e. without fasting) may have produced the results seen in KW. Once again, this is entirely speculative yet leads to a future possible investigation that seems supported by the medical literature referenced in the supplementary analysis section of Section V.

TABLES

Table 1: In order to facilitate comparison with the motivating study, the information in Table 1 is presented in the order and style of Kachelmeier and Williamson (2010). Panel A contains descriptive statistics and the determinants of selection for the present study. ***Bolded and italicized*** means are significantly different between conditions (Quantity-Only vs. Creativity-Weighted) at the 0.01 level; **bolded** means at 0.05; *italicized* at 0.10. Unless otherwise noted, all tests are two-tailed t-tests conservatively assuming unequal variances in the underlying populations. Panel B provides Table 1 from KW to allow for easy comparison of similar measures between studies. Panel C provides p-values comparing KW's measures to those of the present study by condition (Random Assignment vs. Self-Selected). For Panel C, all p-values that are statistically significant at the 0.10 level or higher have been **bolded**.

Table 2: Table 2 presents ANCOVAs with aggregate creativity-weighted productivity (Panels A & B) and total high-creativity output (Panels D & E) as the dependent variables. In both cases the only significant effect is the main effect of contract type. Results demonstrated in Panels A and D are opposite those predicted (for a comparison graphic, please see Figure 4). Opposite the results from KW, creativity-weighted participants significantly outperform their counterparts in the quantity-only contract for both the Number of High-Creativity Puzzles Produced (Panel B; F 4.68; $p < 0.04$, two-tailed) and for Mean Creativity Rating (Panel E; F 5.78; $p < 0.023$, two-tailed). As just mentioned, although neither glucose nor the interaction between glucose and contract type is significant in either case, the pattern of actual results in Table 1 (Panel A and Panel D) surprisingly demonstrates the reverse of my predictions. These findings are discussed in depth in the supplementary analyses portion of Section V of this paper. *Although there is no significant interaction, and I would not do this for a paper being prepared for submission to a journal, I use this opportunity to practice analyzing the simple effects in Panels C and F. Neither panel results in a significant simple main effect.*

Table 1 – Panel A
(Discussion of Table 1 is on the prior page)

TABLE 1
Descriptive Statistics & Determinants of Selection
Panel A - Present Study

	All Self-Selected Contracts		Replenishment Condition (Glucose)		Placebo Condition (Splenda)	
	Mean (sd) {Range}		Mean (sd) {Range}		Mean (sd) {Range}	
	Quantity-Only n = 15	Creativity-Weighted n = 23	Quantity-Only n = 8	Creativity-Weighted n = 11	Quantity-Only n = 7	Creativity-Weighted n = 12
Creativity Potential	4.00 (1.46) {1-7}	4.96 (1.15) {2-7}	4.25 (1.49) {2-7}	5.00 (1.41) {2-7}	3.71 (1.50) {1-5}	4.92 (0.90) {3-6}
Aggregate Creativity-Weighted Productivity	185.67 (76.54) {61-300}	172.98 (70.24) {81-405}	186.42 (67.98) {107-297}	167.39 (52.99) {102-250}	184.81 (90.98) {61-300}	178.11 (85.17) {81-405}
Aggregate Number of High Creativity Puzzles	3.87 (2.85) {1-10}	5.91 (3.32) {1-15}	4.63 (3.38) {1-10}	5.09 (3.02) {1-11}	3.00 (2.00) {1-7}	6.67 (3.52) {2-15}
Quantity	27.33 (11.28) {8-34}	24.21 (10.09) {11-57}	27.13 (10.01) {17-44}	23.82 (7.85) {14-38}	27.57 (13.43) {8-44}	24.58 (12.13) {11-57}
Self-Perceived Task Familiarity	2.66 (1.95) {1-7}	3.48 (1.80) {1-6}	3.00 (2.39) {1-7}	4.18 (1.83) {1-6}	2.29 (1.38) {1-4}	2.83 (1.75) {1-6}

Table 1 – Panel B (Replication of KW table for convenient comparison)

TABLE 1				
Descriptive Statistics				
	Self-Selected Contracts		Randomly Assigned Contracts	
	Mean (Standard Deviation) {Range}		Mean (Standard Deviation) {Range}	
	Quantity-Only	Creativity-Weighted	Quantity-Only	Creativity-Weighted
	n = 24	n = 21	n = 19	n = 26
Creativity Potential	3.92 (1.25) {2-7}	6.14 (1.59) {3-9}	5.26 (1.85) {3-9}	5.15 (1.89) {1-8}
Creativity Rating First Puzzle	4.61 (1.16) {2.11-7.22}	5.09 (1.34) {3.33-7.44}	4.80 (1.22) {2.78-6.78}	4.45 (1.17) {3.00-7.44}
Average Creativity Rating First Eight Puzzles	4.72 (0.59) {3.67-5.82}	5.09 (0.44) {4.42-5.89}	4.98 (0.62) {4.06-6.01}	4.87 (0.51) {3.54-5.83}
Number of High Creativity Puzzles through the Eighth Unit of Production	1.75 (1.36) {0-5}	2.52 (1.21) {0-5}	2.21 (1.62) {0-5}	2.08 (1.29) {0-5}
Aggregate Creativity- Weighted Productivity	113.32 (57.12) {34.56-254.22}	77.31 (29.92) {31.00-141.00}	122.05 (40.98) {57.56-186.78}	83.88 (27.07) {35.00-140.11}
Aggregate Number of High Creativity Puzzles	4.33 (2.84) {0-11}	4.48 (2.06) {1-10}	5.95 (4.01) {0-15}	4.69 (3.06) {1-12}
Quantity	25.13 (15.41) {7-76}	16.85 (6.06) {6-29}	25.74 (10.22) {12-54}	16.85 (5.42) {7-29}
Self-Perceived Task Familiarity	4.92 (2.60) {1-9}	4.86 (2.87) {1-9}	5.26 (2.81) {1-9}	5.19 (3.21) {1-9}

Table 1 – Panel C

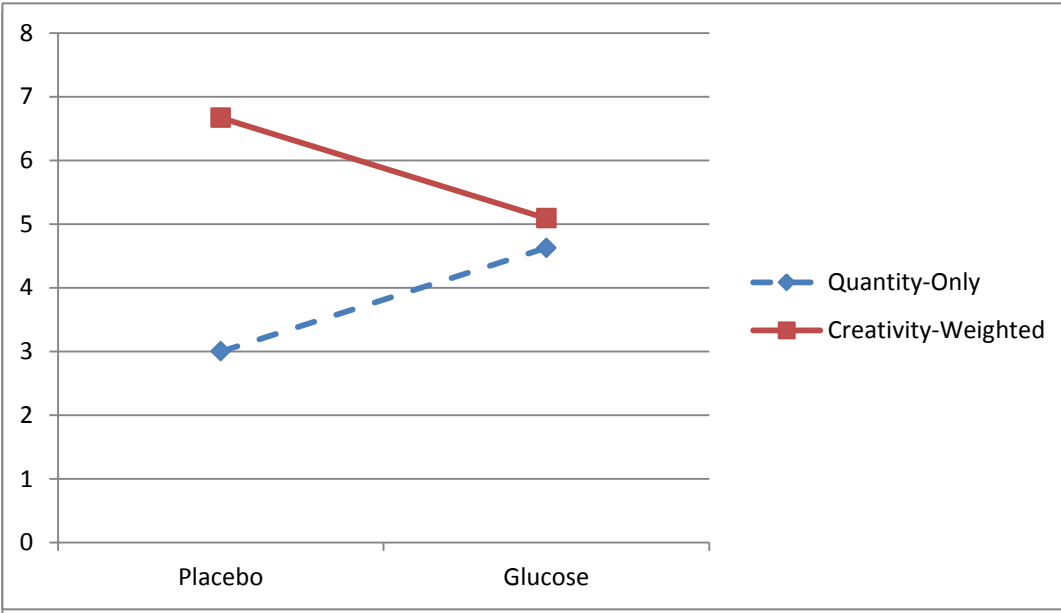
Descriptive Statistics & Determinants of Selection
Panel C - t-Test results comparing present study and KW (2010)

	Comparison of Self-Selected Contracts		Replenishment vs.KW Self-Selected		Placebo vs KW Self-Selected	
	p-values		p-values		p-values	
	<u>Quantity-Only</u>	<u>Creativity-Weighted</u>	<u>Quantity-Only</u>	<u>Creativity-Weighted</u>	<u>Quantity-Only</u>	<u>Creativity-Weighted</u>
Creativity Potential						
vs. KW Self-Selected	0.86	0.01	0.13	0.78	0.00	0.02
vs. KW Random	0.02	0.85	0.00	0.00	0.00	0.58
Aggregate Creativity-Weighted Productivity						
vs. KW Self-Selected	0.01	0.00	0.05	0.00	0.09	0.00
vs. KW Random	0.02	0.00	0.00	0.00	0.13	0.00
Aggregate Number of High Creativity Puzzles						
vs. KW Self-Selected	0.63	0.10	0.36	0.69	0.00	0.07
vs. KW Random	0.05	0.14	0.00	0.00	0.00	0.10
Quantity						
vs. KW Self-Selected	0.62	0.01	0.79	0.02	0.00	0.06
vs. KW Random	0.73	0.01	0.00	0.00	0.00	0.06
Self-Perceived Task Familiarity						
vs. KW Self-Selected	0.01	0.07	0.06	0.23	0.00	0.03
vs. KW Random	0.00	0.02	0.01	0.00	0.00	0.01

TABLE 2: Hypothesis Tests
(Discussion of Table 2 is on the heading page for the Tables Section)

**ANCOVA for the Effect of Contract Type and Treatment (Glucose)
on the Number of High-Creativity Puzzles Produced**

Panel A: Mean number of High-Creativity Puzzles Produced



Panel B: ANCOVA on number of High-Creativity Puzzles Produced

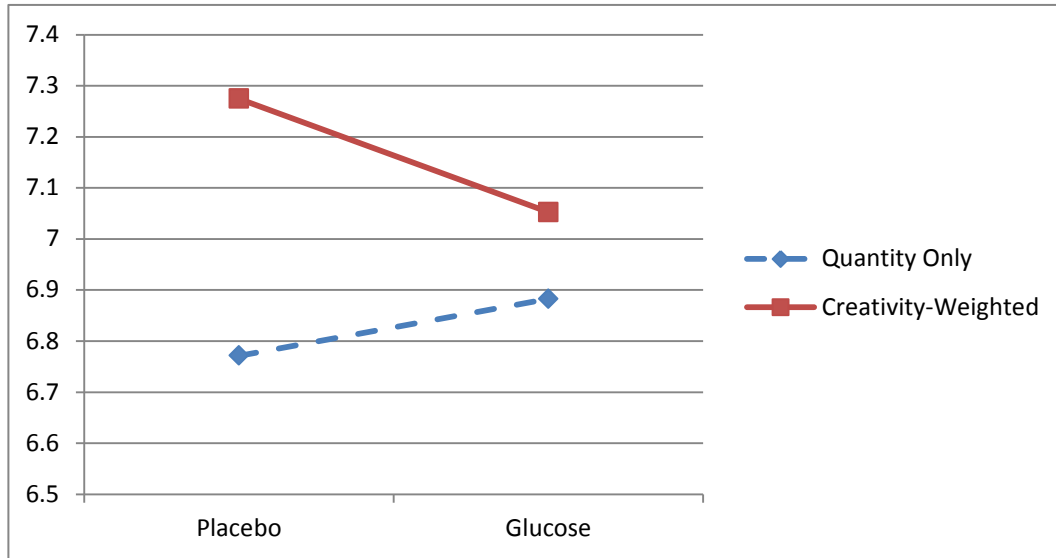
FACTOR	df	MS	F	p-value (two tailed)
Model	9	19.49	2.5	0.03
Contract Type	1	36.44	4.68	0.039
Cognitive Replenishment (Glucose)	1	2.598	0.33	0.568
Contract Type X Glucose	1	17.19	2.21	0.148
Task Familiarity (covariate)	6	18.88	2.42	0.051
Error	28	7.79		

Panel C: Simple Effects

FACTOR	df	MS	F	p-value (two tailed)
Effect of Glucose within Creativity-Weighted	1	14.25	1.829	0.187
Effect of Glucose within Quantity-Only	1	9.85	1.264	0.27

TABLE 2 – Continued
(Discussion of Table 2 is on the heading page for the Tables Section)
ANCOVA for the Effect of Contract Type and Treatment (Glucose)
on Mean Creativity Rating

Panel D: Mean Rating of All Puzzles Produced



Panel E: ANCOVA on Mean Rating of Puzzles Produced

FACTOR	df	MS	F	p-value (two tailed)
Model	9	0.312	2.64	0.0236
Contract Type	1	0.683	5.78	0.023
Cognitive Replenishment (Glucose)	1	0.065	0.55	0.4639
Contract Type X Glucose	1	0.209	1.77	0.1939
Task Familiarity (covariate)	6	0.240	2.04	0.0941
Error	28	0.118		

Panel F: Simple Effects

FACTOR	df	MS	F	p-value (two tailed)
Effect of Glucose within Creativity-Weighted	1	0.1179	1	0.326
Effect of Glucose within Quantity-Only	1	0.046	0.39	0.538