Do Vesting Requirements Increase the Incentive Effects of Stock Compensation for Rank-and-File Employees?

by

Jeffrey Douglas Clark

B.S. Accounting, Brigham Young University, 2014

MAcc, Accountancy, Brigham Young University, 2014

Submitted to the Graduate Faculty of

Joseph M. Katz Graduate School of Business in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Business Administration

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

JOSEPH M. KATZ GRADUATE SCHOOL OF BUSINESS

This dissertation was presented

by

Jeffrey Douglas Clark

It was defended on

April 22, 2019

and approved by

Jongwoon (Willie) Choi
Associate Professor of Accounting
University of Wisconsin-Madison

John H. Evans III
Katz Alumni Professor of Accounting and Area Director for Accounting
University of Pittsburgh

Vicky B. Hoffman
Professor of Business Administration and Ben L. Fryrear Faculty Fellow
University of Pittsburgh

Dissertation Advisor: Donald V. Moser
Professor of Business Administration and Lou and Myra G. Mervis Chair
University of Pittsburgh

Do Vesting Requirements Increase the Incentive Effects of Stock Compensation for Rank-and-File Employees?

Jeffrey Douglas Clark, PhD

University of Pittsburgh, 2019

My dissertation consists of three chapters. The first chapter's primary purpose is to introduce readers to my dissertation. In it, I note the importance of research on motivating employees to provide effort. I also introduce my specific topic, how vesting requirements moderate the effectiveness of stock-based compensation in motivating employee effort, and discuss challenges that researchers confront when studying employee effort.

In the second chapter, my co-authors, Willie Choi and Adam Presslee, and I identify seven factors researchers should consider when designing a real-effort task to ensure a strong, positive, and consistent link between participants' effort intensity and their performance. With these factors in mind, we design an experiment to test for incentive effects on effort and performance using three effort-intensive tasks: the decode task, the letter search task, and the slider task. Contrary to our expectation, we find significant variation across tasks in our ability to detect incentive effects and limit the effects of the seven factors, with the strongest evidence coming from the slider task. Our list of factors helps researchers design effort-intensive tasks that allow them to conduct a more effective test of theory. Following this conclusion, I use the slider task to effectively test the theories I use in the third chapter of this dissertation.

In the third chapter, I provide evidence that firms commonly compensate rank-and-file employees with restricted stock that has vesting requirements, despite economic arguments that it is a less effective incentive than cash. Using an experiment, I investigate whether restricted stock

compensation can motivate greater employee effort than other economically equivalent contracts. First, I demonstrate individuals view restricted stock compensation as a penalty contract. Second, I leverage research on framing and the endowment effect to predict a penalty-framed stock contract (representing restricted stock) will be valued more highly and motivate greater effort than an economically equivalent penalty-framed cash contract or an economically equivalent bonus-framed stock or cash contract. The results of my experiment are consistent with this prediction. That is, I find effort is greatest under a contract with the features of restricted stock with vesting requirements, which helps explain the prevalence of such compensation arrangements for rank-and-file employees.

Table of Contents

List of Tables

List of Figures

Preface

I thank the past and present accounting faculty at the University of Pittsburgh. I had the privilege of taking seminars from almost every faculty member, and in addition to being great teachers, they are great friends and mentors. In particular, I thank my dissertation committee members, Willie Choi, Harry Evans, Vicky Hoffman, and my dissertation chair, Don Moser.

Willie Choi provided my first opportunity to work on a co-authored research project and always showed me great respect even though I had a lot to learn. He taught me to enjoy research by working hard, while not forgetting to celebrate milestones early and often. Willie also took time to mentor me in other areas, such as learning how to properly do a journal review, to help me prepare for my career. As a result, I became more confident in my ability to be successful, but Willie made sure I did not get too big of an ego by humbling me on the basketball court several times a week.

Harry Evans was the first person to make me feel welcome at Pitt. Soon after I arrived in Pittsburgh, he asked for my address so he could bike to my home and welcome my family. He was undeterred when he found out this would be at least a two-hour bike ride, and this is representative of how willing Harry is to spend his time helping others (and how much he likes biking!) Harry also allowed me to be his teaching assistant for his MBA accounting course, which played a large role in helping me prepare to teach my own courses.

Vicky Hoffman was my first point of contact at Pitt and was instrumental in my decision to accept an offer here. She convinced me the faculty would help me achieve my goals, and she followed up on that promise by being one of my most consistent sources of support during the last five years. I appreciate her willingness to allow me to stop by her office unannounced as she

supported me not only in completing my dissertation, but also with helpful advice throughout my search for a job.

Don Moser has been an incredible mentor since the day I started at Pitt, and I am forever indebted to him for all the work he did to help me graduate. During the course of the doctoral program, there were many late nights and Don was often working right there with me. There were countless times he stayed up until the wee hours of the night to ensure that, in addition to his numerous other responsibilities, he made time for me so I was never lacking in support. Don's support was especially key in helping me complete my dissertation. He had the insight to suggest I apply my proposed theories to a broader setting, and spent hours upon hours editing my document to make sure others could understand and appreciate my contributions. Finally, on top of all the professional help Don provided, he always found ways to express that he cared about me as an individual and not just as a researcher. This was done in various ways, ranging from inviting me to attend sporting events with him to sending thoughtful emails checking in on how my kids were doing.

These individuals were so important to me during the past five years and I am very grateful for their positive contributions to my experience as a doctoral student. I would be remiss if I did not also single out two more faculty members for their contributions. Adam Presslee is full of positivity and believed in me more than I believed in myself. He was often the first person I'd approach when I ran into a problem and was my go-to stats guru for complicated issues. Adam's willingness to teach a seminar just for me during my first year was also an essential part of my development. Not only did it introduce me to many interesting topics, it spurred me to, with Adam's help, develop the idea that eventually became my job market paper. Nicole Cade was a massive support during my search for a job. Her advice helped guide me every step of the way, and her taking the time to listen to my practice presentations and read through my application materials led to significant improvements.

I am also indebted to numerous others. I thank my fellow PhD students (Eric Chan, Michele Frank, Duanping Hong, Yangyang Fan, Jordan Bable, Melinda Ford, Brian Knox, Jared Smith, Yue Zhang, Conor Brown, Kristin Stack, Will Docimo, Liz Connors, Nathan Mecham, and Benda Yin) for great support and fun conversations over the years. These students are lifelong friends and I wish space permitted me to talk of their individual contributions. I also thank Carrie Woods, Chris Fedor, Chris Gursky, Natalia Fenton, and Rachael McAleer for solving all my administrative problems and keeping the doctoral office stocked with chocolate.

Finally, I thank my family. My wonderful wife, Madi, was my rock during the doctoral program. She took on an outsized role in our family and continually did more than her fair share of the familial work to allow me time for work. On top of that, she still found ways to directly contribute to my completion of program requirements, using her superior calculus skills to help me through microeconomics courses and her copy-editing skills to always give my documents a first read. I also thank my children, Ellie and Owen, for giving me a pick-me-up every time I arrived home and for their unconditional love. Last, I thank my parents for constant encouragement and many entertaining conversations during my commutes to and from school.

1.0 Introduction

This dissertation consists of three chapters. Chapter 1.0 introduces my research topic and provides a brief overview of Chapters 2.0 and 3.0. Chapter 2.0 presents an analysis of various experimental tasks that may be used to study employee effort in a laboratory setting. The dissertation's primary study in Chapter 3.0 uses the task design supported by Chapter 2.0 to analyze how vesting requirements moderate the effects of stock compensation on effort in rank-and-file contracts.

Understanding how to motivate employees to provide effort is of great interest to practitioners and accounting researchers (Bonner and Sprinkle 2002; Merchant and Van der Stede 2017). In my dissertation, I aim to inform the reader on the effectiveness of vesting requirements in stock compensation in motivating employees to provide effort. To study how compensation contracts affect employee effort, many researchers conduct experiments that use a choice-effort task in which participants choose an effort level from a menu of effort levels, each of which carries with it a specific monetary cost to the employee (e.g., Kuang and Moser 2009). In these choice-effort tasks, the monetary cost is increasing in the effort level to reflect the standard assumption of effort aversion. An alternative approach, used by experimental researchers to measure participants' effort, is to use a real-effort task, in which participants perform an actual task for a set amount of time and earn rewards based on their performance on that task.

One major advantage of using a choice-effort task instead of a real-effort task is that a choice-effort task allows researchers to capture effort directly rather than rely on task performance, which is a noisy proxy for effort. Task performance is a noisy proxy of effort because observed differences in performance may be caused by factors other than differences in effort, such as differences in individual intrinsic motivation or ability. These confounding factors make it difficult for researchers to draw conclusions from studies that use real-effort tasks.

However, real-effort tasks are necessary to properly test some theories, including one of the theories that I rely on in my dissertation. Specifically, I rely on the endowment effect which shows that ownership of an item causes an individual to value it more highly. However, an exception to the endowment effect is that the valuation of cash is not affected by ownership because it is difficult to mentally revalue. Therefore, I predict that the endowment effect will differentially affect owners of stock and owners of cash. If I were to use a choice-effort task, participants would have to choose an effort level, with its associated monetary cost. This would force participants to jointly evaluate the stock and the monetary cost of effort required to obtain it. Thus, my participants would likely mentally view any stock compensation as simply its cash equivalent such that the endowment effect's influence is artificially eliminated in my experiment.

To summarize, there are theoretical reasons why I need to use a real-effort task in my study, but such tasks rely on a potentially noisy proxy of effort. Therefore, before I could proceed to my main research question addressed in Chapter 3.0, I needed to design a real-effort task in which task performance is sensitive to changes in effort and is minimally affected by other factors that create noise.

In Chapter 2.0, along with my co-authors, Willie Choi and Adam Presslee, I identify seven factors researchers should consider when designing a real-effort task to ensure a strong and consistent link between participants' effort and performance. With these factors in mind, we design an experiment to test for incentive effects on effort and performance using versions of three different real-effort tasks. We find significant variation across tasks in our ability to detect incentive effects. The slider task, as described in Chapter 2.0, was most effective at capturing incentive effects and mitigating the effects of factors other than effort. Therefore, I use the slider task in my main study on how vesting requirements affect the incentive effects of stock compensation.

I address my main research question in Chapter 3.0, where I provide evidence that firms commonly compensate rank-and-file employees with restricted stock that has vesting requirements, despite economic arguments that restricted stock is a less effective incentive than cash. Using an experiment, I predict and find that participants value stock more highly in contracts with the features of vesting requirements. Further, because of these higher valuations in stock contracts with vesting requirements, participants provide more effort under these contracts than other economically equivalent stock and cash compensation contracts. These motivational benefits offer a possible explanation for why restricted stock contracts with vesting requirements are so common in rank-and-file compensation contracts.

2.0 Testing the Effect of Incentives on Effort Intensity Using Real-Effort Tasks

2.1 Introduction

Understanding how incentive pay affects employee effort intensity and performance is a topic of great interest to accounting practitioners and researchers (Bonner and Sprinkle 2002; Datar and Rajan 2014). Many researchers investigate this topic using laboratory experiments in which participants perform an effort-intensive task for a set amount of time and earn rewards based on their task performance (Bonner, Hastie, Sprinkle, and Young 2000).[1] Since participants' task performance serves as a proxy for their effort intensity, it is critical for researchers to design the task such that task performance is sensitive to changes in effort intensity and is minimally affected by other factors. The purpose of our study is two-fold. First, we identify seven factors researchers should consider when designing a real-effort task to ensure a strong, positive, and consistent link between participants' effort intensity and their performance. Second, with these factors in mind, we conduct an experiment in which we develop versions of several commonly used effort-intensive tasks and test their suitability for detecting variations in participants' effort intensity.

The seven factors are: (1) intrinsic motivation; (2) task-specific skill; (3) task strategies beyond exerting effort; (4) fine unit of performance; (5) within-round task experience across participants; (6) performance trends over time; and (7) task difficulty across rounds. In our experiment, we apply our insights regarding these factors to the design of three effort-intensive

[1] In addition to using a real-effort task, researchers can use a choice-effort task to study effort. In Appendix A, we discuss the relative merits of using real-effort versus choice-effort tasks.

tasks commonly used by prior accounting research. First, we design a version of the decode task, in which participants decode numbers into letters using a decoding key (Chow 1983; Fisher, Maines, Peffer, and Sprinkle 2002; Church, Libby, and Zhang 2008). Second, we design a version of the letter search task, in which participants count the number of times a specific letter appears within a grid (Sprinkle, Williamson, and Upton 2008; Kachelmeier, Thornock, and Williamson 2016). Finally, we design a version of the slider task, in which participants drag a slider box to a specific point on a slider bar (Gill and Prowse 2012, 2015; Chan 2017).

Since the goal of our experiment is to evaluate experimental task designs and not to test new theory, we employ a simple experimental design in which participants perform one of the tasks and we manipulate (between participants) whether participants receive piece-rate pay or fixed pay. Since both economic and psychology theory predict individuals will work harder under incentive pay (see Bonner and Sprinkle 2002 for review), we expect performance will be greater under piece-rate pay than under fixed pay (i.e., observe incentive effects). Thus, in our experiment, we evaluate the effectiveness of our task designs by testing whether we are able to detect incentive effects with each task and mitigate problematic effects related to the seven factors above. Surprisingly, despite our best attempts, we find significant variation across tasks in our ability to detect incentive effects and to limit the effects of the seven factors above. Ultimately, we find the slider task is the most effective task.

Our study contributes to research by helping researchers who are considering the use of effort-intensive tasks. Our list of factors will help researchers conduct a more effective test of theory using effort-intensive tasks by offering guidance on establishing a strong, positive, and consistent link between effort intensity and task performance. In addition, we highlight the difficulty of using such tasks to detect incentive effects in a laboratory experiment, and our analysis of design factors offers some insights into why this might be. Finally, we provide evidence the slider task may be the most effective task for researchers to use when studying the effect of incentives on effort intensity.

The remainder of the paper is organized as follows. Section 2.2 describes design factors for researchers to consider when designing a real-effort task to capture effort intensity. Section 2.3 describes the experimental design, and Section 2.4 details our results. We offer some concluding remarks in Section 2.5.

2.2 Important Design Considerations

In order to strengthen the inferences regarding the effects of incentive pay on effort intensity using real-effort tasks, it is important for researchers to minimize the effects of other factors that make it difficult to use task performance as a proxy for effort intensity.[2] Accounting for these factors helps to reduce noise and bias when conducting tests of the effects of incentive pay on effort.[3] In this section, we discuss seven important factors researchers should account for when designing an effort-intensive task: (1) intrinsic motivation; (2) task-specific skill; (3) task strategies beyond exerting effort; (4) fine unit of performance; (5) within-round task experience across participants; (6) performance trends over time; (7) task difficulty across rounds. The first five factors are those we believe researchers should consider whenever they use an effort-

[2] We acknowledge researchers can account for these factors in two, non-mutually exclusive, ways: (1) design the task in ways to minimize the existence of these factors, and (2) measure and statistically control for the effects of these factors. We focus on (1) because it is likely a more effective approach for addressing the factors we discuss in this section.

[3] Bonner and Sprinkle (2002) offer a broader, conceptual framework of the relationships among compensation schemes, effort, and performance, and we encourage interested readers to read their paper.

intensive task, while the last two factors are more relevant for studies using an experimental design in which participants perform an effort-intensive task over multiple rounds.

2.2.1 Intrinsic Motivation

Individuals are intrinsically motivated when an action provides them with inherent rewards that are not dependent on receiving any external reward (Deci, Koestner, and Ryan 1999). There are at least two reasons participants may be intrinsically motivated to exert effort. First, participants could view a task as "game-like" or otherwise enjoyable in nature (Fessler 2003; Farrell, Grenier, and Leiby 2017). Second, participants could view performing well on the task as being important to their self-concept (Tafkov 2013).

Greater levels of intrinsic motivation can attenuate the effectiveness of incentive pay on effort (Fessler 2003; Farrell et al. 2017).[4] Specifically, participants who are more intrinsically motivated to exert effort are less likely to be effort-averse such that they exert high effort regardless of incentive condition, which creates an upper bound issue in terms of effort and performance. Notably, randomization of participants to treatment condition does not fully address this upper bound issue because high levels of intrinsic motivation in all conditions reduces the differences in effort between conditions that can otherwise be explained by the treatment.

[4] Contrary to claims by prominent commentators (e.g., Kohn 1993; Pink 2009), there is compelling empirical evidence that incentive pay does not crowd out intrinsic motivation in work settings (Hossain and Li 2014; Shaw and Gupta 2015).

2.2.2 Task-Specific Skill

Differences in participants' task-specific skill can have a substantive impact on task performance. Two participants exerting the same amount of effort may have different performance outcomes if one is more skilled than the other, especially if task performance depends relatively more on participants' skill than on their effort. For example, if the real-effort task involves participants answering trivia questions, then differences in performance are more likely to reflect differences in participants' skill (knowledge) than differences in participants' effort because "trying harder" (e.g., spending more time thinking about the question) will likely not change the fact that an answer is not known (Libby and Luft 1993; Bonner and Sprinkle 2002).

In addition, in order for task performance to be a good proxy for effort, all participants must have the requisite skills to complete the task, even if task-specific skill has only a minimal effect on performance (Bonner et al. 2000). If participants lack the requisite skills and are aware of that deficiency, then incentive pay is less likely to affect effort because participants will assess the relationship between effort and performance to be weak. Indeed, as the difference between the participants' skills and the requisite skills to perform the task increases, research suggests task performance depends more on factors such as whether participants identify a suitable strategy to perform the task than on their effort (Locke and Latham 1990).

2.2.3 Task Strategies Beyond Exerting Effort

When using an effort-intensive task, it is important participants exert effort only in the manner the researchers intend. If participants are able to perform the task in a way that researchers find undesirable, then participants' performance on the task may not be a good proxy for their effort intensity. For example, in Gill and Prowse's (2012, 2015) version of the slider task, participants are presented with a grid of sliders with endpoint positions of 0 and 100, each with a slider box that moves along the slider bar in increments of 1. Each slider box is initially positioned at the 0-position, and the objective of the task is to move the slider box to the 50-position *using the computer mouse*. Thus, the intent is for participants to click and drag the slider box for each slider, and not move the slider box using other means. To ensure this, the authors disabled the computer keyboard to prevent "participants from using the arrow keys to position the sliders" (2012, 473).

On some real-effort tasks, guessing can also act as an undesirable, alternative approach to performing the task. For example, some studies use a letter search task in which participants count the number of times a given search letter appears in a grid of letters. In this task, participants can expend effort to carefully scan the grid to find the search letter (presumably, what researchers desire regarding effort intensity), or participants can submit a guess after doing a quick scan and then make minor modifications to their initial guess, if necessary. If guessing is not deterred, e.g., through a time penalty or a reduction in performance for incorrect answer submissions, then task performance could be attributable to luck rather than to effort intensity.

2.2.4 Fine Unit of Performance

Participants' task performance is a better proxy of their effort intensity when each unit of performance is (sufficiently) fine. For example, many studies use Chow's (1983) decoding task in which participants translate numbers into letters using a provided decoding key. However, in some studies, translating a single number into a letter constitutes one unit of performance, while in other studies, participants must translate a set of two or more numbers into letters to generate one unit of performance. Thus, the latter design requires more "components" to be completed in order for participants to generate one unit of performance.

When the number of required components increases, the likelihood that participants will have a partially completed unit of performance at the end of a round also increases. Such cases are problematic because they make it difficult for researchers to clearly distinguish these cases from those in which participants did not have any partially completed units. This issue is particularly acute when one unit of performance constitutes a significant percentage of total performance in a round.

Finally, there can still be inferential problems even if there are no partially completed units at the end of the round. Specifically, when multiple components are required to complete one unit of performance, participants may stop exerting effort near the end of a round because they believe they do not have time to finish another unit before time expires in the round. Such cases bias against detecting incentive effects using an effort-intensive task.

2.2.5 Within-Round Task Experience Across Participants

To strengthen the inferences about effort intensity using a real-effort task, it is important to ensure participants experience the task in an identical manner, i.e., to create a similar within-round task experience across participants. Otherwise, the effects of incentive pay on effort are confounded with differences in participants' experiences with the task. For example, when using Chow's (1983) decoding task, researchers should ensure that in a given round, all participants see the same decoding key and decode the same sequence of numbers during the round. This ensures the difficulty level of the task in a given round is identical for all participants, and allows researchers to draw stronger inferences about the effects of incentive pay on effort intensity.

2.2.6 Performance Trends Over Time

When studies involve a setting with multiple rounds, participants gain substantial practice – and thus, greater comfort and familiarity with the task – through repeated exposure. As such, participants' performance may improve over time simply due to practice (i.e., improved skill) and not due to greater effort intensity. Such learning effects are particularly troublesome in studies using a pretest-posttest experimental design. Notably, trying to measure or statistically control for such learning effects by including a practice round may not always be effective. For example, Gill and Prowse (2015) report participants in their experiment demonstrate significant non-linear learning effects even though participants perform a real-effort task in two 2-minute practice rounds before the main performance rounds.

The learning effects described above capture an upward trend in performance over time. However, a downward trend in performance over time may also arise, as participants performing a task over multiple rounds may start to feel fatigued. This occurs due to deficits in attention, working memory and action control (Lorist et al. 2009; Möckel, Beste and Wascher 2015).

2.2.7 Task Difficulty Across Rounds

In addition to learning effects and fatigue, researchers using an effort-intensive task in an experiment with multiple rounds should be cognizant of variations in performance over time due to differences in task difficulty across rounds. For example, suppose the task involves solving multiplication problems in which two 2-digit numbers are multiplied together, and participants must solve the problems without the aid of a calculator or pen and pencil. For this task, the difficulty of the multiplication problems can vary greatly, from relatively easy problems (e.g., 10 x 10) to relatively difficult problems (e.g., 74 x 68). If the problem sets in every round are not consistent in the relative mix of easy and difficult problems, then variations in task difficulty across rounds can arise. As a result, participants' performance becomes a noisier proxy for their effort intensity because a participant may have performed differently in one round compared to another round simply because the set of multiplication problems in one round was easier or harder, and not because the participant changed the amount of effort exerted on the task.

2.3 Method

2.3.1 Overview

We complement the discussion in Section 2.2 with an experiment using various effort-intensive tasks from recent accounting research. The goal of our experiment is to apply our insights from Section 2.2 and design a version of each task we believe is effective for capturing effort intensity. Since our focus is on evaluating task design and not on testing new theory, we employ a simple experimental design in which participants perform one of the tasks and we manipulate (between participants) the compensation scheme (incentive pay versus fixed pay). As noted earlier, we focus on the effects of incentive pay on effort intensity because this is a topic of interest to accounting practitioners and researchers.

A large body of research in management accounting and elsewhere find incentive pay positively affects effort and task performance (e.g., Bonner and Sprinkle 2002; Luft and Shields 2003; Shaw and Gupta 2015; Herschung, Mahlendorf, and Weber 2017). Thus, to evaluate how effectively our designed tasks capture variations in participants' effort intensity, we test whether we are able to detect incentive effects with each task. In the following subsections, we first describe the task selection process, followed by the compensation scheme manipulation. Then, we describe the participants and the session logistics. Finally, we describe the global and task-specific features we employed in our experiment.

2.3.2 Task Selection

To determine which tasks to include in our experiment, we generated an initial list of real-effort tasks by reviewing experimental research on effort published or forthcoming in eight accounting journals (TAR, JAR, JAE, CAR, AOS, RAST, JMAR, BRIA) from 2000 – 2018 (See Appendix B). As Appendix B shows, the vast majority of prior research using real-effort tasks focuses on effort intensity, although some studies do examine other dimensions of effort. Since we focus on effort intensity as well, we excluded real-effort tasks that capture multiple dimensions of effort or dimensions of effort other than effort intensity. For example, we excluded Sprinkle's (2000) production task and Tafkov's (2013) multiplication task because these tasks capture effort duration rather than effort intensity. Similarly, we excluded Farrell, Kadous, and Towry's (2008, 2012) sandwich-making task and Kachelmeier, Reichert, and Williamson's (2008) creativity task because these tasks capture effort direction (e.g., quality) in addition to effort intensity (e.g., quantity). After excluding these types of tasks, we settled on three effort-intensive tasks: the decode task, the letter search task, and the slider task.

2.3.3 Compensation Scheme Manipulation

To assess whether the tasks are suitable for detecting variations in participants' effort intensity, we administered two different compensation contracts for each task: a piece-rate pay contract and a fixed pay contract. For the decode task, we administered a piece-rate of $0.35 per unit of performance. For the letter search task, we administered a piece-rate of $0.10 per unit of performance. For the slider task, we administered a piece-rate of $0.30 per unit of performance. We chose these piece-rate amounts based on pilot-testing, such that participants would earn around $10, on average.[5] For each task, we administered the piece-rate pay contract condition first, and then calculated the average pay participants earned under the piece-rate pay contract for each task. We used these averages to create the fixed pay contract for each task (e.g., Webb, Williamson, and Zhang 2013). Thus, within each task in our experiment, we held constant the average compensation earned under each contract. For the decode task, the fixed pay contract offered $9.35. For the letter search task, the fixed pay contract offered $13.50. For the slider task, the fixed pay contract offered $11.00.[6]

[5] In our pilot tests, nine doctoral students completed two rounds of each task, with each round lasting two minutes. They did not earn any compensation for their performance, but were asked to perform as best as they could on the tasks.

[6] For the decode task, average piece-rate pay is significantly less than $10 ($t = 2.54$, $p = 0.01$). For both the letter search task ($t = 15.71$, $p < 0.01$) and the slider task ($t = 3.88$, $p < 0.01$), average piece-rate pay is significantly greater than $10. Importantly, our focus is to compare task performance between the two compensation schemes within each task, and not across tasks. Thus, any deviations of average compensation from $10 are unlikely to pose inferential difficulties in our analysis.

2.3.4 Participants and Session Logistics

One hundred eighty-nine individuals recruited from an experimental economics participant pool completed our experiment. Participants were 20.3 years old, on average, and approximately 57 percent were female. Age and gender are uncorrelated with performance, with one exception. For the slider task, the correlation between age and performance is negative and statistically significant ($r = -0.30$, $p = 0.02$). Our inferences are unchanged if we control for age in our analyses. Therefore, the main analyses reported in the paper do not control for age, and we do not discuss age in the remainder of the paper.

We administered the experiment in an experimental economics laboratory, and all participants completed the experiment using a computer. Conducting our experiment in this lab enabled all participants to complete the experiment with identical hardware (monitors, keyboards and mice) and software (internet browser, java version). All computers included Windows 7 Enterprise with 19-inch monitors calibrated to a 1440 x 900-pixel resolution. We used Google Chrome as the browser, and all computers included identical USB mice and keyboards. So, any differences we observe in task performance do not reflect differences in hardware and software capabilities.

We administered one cell (unique contract and task combination) per session, and we conducted two sessions for each cell. Our inferences are unchanged if we include session as a variable in our analyses. Thus, the main analyses reported in the paper do not control for session, and we do not discuss this variable in the remainder of the paper.

At the start of each session, participants received initial instructions describing the task, and then completed a two-minute practice round. Next, participants received additional instructions detailing the compensation scheme, and completed an instructions quiz. We required participants to answer each quiz question correctly before proceeding to the next stage of the experiment. After successfully completing the quiz, participants performed the task for eight two-minute main performance rounds. In both the practice round and the main performance rounds, the computer screen displayed a timer indicating the time remaining in the round. During each round, participants received real-time performance feedback, and they also received feedback at the end of each round regarding their performance and pay for that round. Then, participants completed a post-experimental questionnaire and received payment. We randomly selected one performance round for payment, and paid participants in private before they exited the lab. Each session lasted about 30 minutes.

2.3.5 Task Features

2.3.5.1 Global Features

Before describing the task-specific design choices, we highlight global design choices applied to the tasks; see Table 1 for a list of global and task-specific features and how they relate to the factors discussed in Section 2.2, and Appendix C for screenshots of each task. First, as noted above, participants perform the task for nine total rounds (a two-minute practice round and eight two-minute main performance rounds). Including a practice round should reduce learning effects in the main performance rounds, and should minimize the effects of differences in task-specific skill by providing participants with an opportunity to familiarize themselves with the mechanics of performing the task. Including multiple main performance rounds allows us to assess whether learning effects still emerge even after a practice round. Notably, we did not compensate participants for the practice round because doing so would introduce a new factor, i.e., whether participants experience a switch in compensation, and the effects of our compensation scheme manipulation could be confounded with this new factor. Such a confound would be especially problematic if participants earned piece-rate pay in the practice round but subsequently experienced a fixed pay contract in the main performance rounds.

Second, each round lasts two minutes. This duration is long enough to detect variations in performance when there is a fine unit of performance. But, it is also short enough to limit the effects of working memory (a potential task-specific skill) and other, undesirable phenomena such as fatigue.

Third, for the decode and letter search tasks, we assessed a three-second time penalty for incorrect answers to discourage participants from pursuing a guessing strategy. Since the nature of the slider task precludes guessing, we did not assess a time penalty for that task.

Fourth, we attempt to minimize participants' intrinsic motivation through several design features. We do not assign any absolute (e.g., individual performance goals) or relative performance standards (e.g., relative performance information), although we do provide real-time individual performance feedback in each round. Providing either absolute or relative performance standards could increase participants' intrinsic motivation to perform the task. We use an abstract (context-free) setting to further reduce intrinsic motivation to perform the task, as prior research suggests individuals may derive utility from the act of working (Webb et al. 2013). In addition, our focus on a single dimension of effort, effort intensity, helps reduce intrinsic motivation because real-effort tasks with multiple dimensions (or having participants perform multiple real-effort tasks) offers participants a greater sense of autonomy and control, both of which are positively related to intrinsic motivation (Deci et al. 1999). Also, the real-effort tasks we examine in our experiment are likely to have a weak link to participants' self-concept, which helps reduce intrinsic motivation (e.g., Tafkov 2013).

Fifth, all three real-effort tasks in our experiment are simple production tasks which are not overly complex (see Bonner et al.'s (2000) taxonomy of tasks in terms of their complexity). Thus, we expect all participants will have the requisite skill to perform any of the tasks.

2.3.5.2 Task 1: Decode Task

In the decode task, participants translate numbers into letters using a provided decoding key (Chow 1983). In our experiment, participants are shown one three-digit number at a time, and they translate the number into the appropriate letter using a decoding key, which is also shown on the bottom of the screen.

When designing this task, we made four important design choices. First, as noted earlier, each round lasts two minutes, as this duration is short enough to limit the effects of working memory, a task-specific skill, on performance. By limiting the effects of working memory, the relatively short duration of each round is also consequential for the decode task because it should also discourage participants from trying to memorize portions of the decoding key in order to improve their task performance, which relates to limiting task strategies beyond exerting effort.

Second, we defined a unit of performance as one correctly decoded three-digit number. Some prior studies using the decode task define a unit of performance as the successful decoding of multiple numbers (e.g., Kelly, Webb, and Vance 2015; Kelly and Presslee 2017). However, we chose to define a unit of performance as a single correct decode because it is a finer unit of measurement. As noted earlier, if participants are required to decode multiple numbers successfully to generate one unit of performance, it would not be possible to distinguish participants who partially complete a unit of performance at the end of a round from those who make no progress on their final unit.

Table 1: Task Features and Connections to Factors Discussed in Section 2.2

Panel A: Global Task Features

	Intrinsic Motivation	Task-specific Skill	Task Strategies Beyond Exerting Effort	Fine Unit of Performance	Within-Round Task Experience Across Participants	Performance Trends Over Time	Task Difficulty Across Rounds
Global Task Features							
1) Practice round prior to the performance rounds		✓				✓	
2) Rounds were two minutes in length		✓		✓		✓	
3) 3 second penalty for errors			✓				
4) No absolute or relative performance standards	✓						
5) Abstract, context-free setting	✓						
6) Single dimension of effort	✓						
7) Weak link between domain and self-concept	✓						
8) All tasks are simple production tasks (Bonner et al. 2000)		✓					

Panel B: Decode Task

	Intrinsic Motivation	Task-specific Skill	Task Strategies Beyond Exerting Effort	Fine Unit of Performance	Within-Round Task Experience Across Participants	Performance Trends Over Time	Task Difficulty Across Rounds
Decode Task Features							
1) Rounds were two minutes in length		✓	✓				
2) One decode = one performance unit				✓			
3) Change decoding key at the end of every round		✓	✓			✓	
4) All participants use the same decoding key in a given round					✓		
5) All participants see the same sequence of numbers to decode in a given round					✓		

Panel C: Letter Search Task

	Intrinsic Motivation	Task-specific Skill	Task Strategies Beyond Exerting Effort	Fine Unit of Performance	Within-Round Task Experience Across Participants	Performance Trends Over Time	Task Difficulty Across Rounds
Letter Search Task							
1) The pool of potential search letters is two			✓				
2) The size of the letter search grid is one cell		✓	✓	✓			✓
3) One completed letter search = one performance unit				✓			

Panel D: Slider Task

	Intrinsic Motivation	Task-specific Skill	Task Strategies Beyond Exerting Effort	Fine Unit of Performance	Within-Round Task Experience Across Participants	Performance Trends Over Time	Task Difficulty Across rounds
Slider Task Features							
1) Staggered formation of adjacent sliders			✓				
2) Change the alignment of sliders every round		✓				✓	
3) All sliders in each round are visible on the screen simultaneously		✓				✓	✓
4) Force participants to drag slider boxes			✓			✓	
5) One slider = one performance unit				✓			
6) All participants see the same staggered formation of sliders in a given round					✓		

This table presents the mapping of global and task-specific features to the factors discussed in Section 2.2. For each global or task-specific feature, we highlight the relevant factors with a ✓. See Section 2.3.5 for a detailed discussion of the mapping of global and task-specific features to the factors discussed in Section 2.2.

Third, we provided participants with a new decoding key in each round (e.g., Kelly and Presslee 2017). This design choice helps minimize the effects of task-specific skill by limiting the effect of participants' working memory on performance. If the same decoding key is used in every round, then some individuals may be able to remember portions of the decoding key, which allows participants to perform better on the task. Consequently, differences in performance may reflect differences in ability and not in effort intensity. Moreover, providing participants with a new decoding key in each round discourages participants from expending effort to memorize the decoding key, which minimizes the potential for participants to invest in task strategies other than exerting effort. In addition, by limiting the effect of participants' working memory on performance, providing participants with a new decoding key in each round also increases the likelihood participants will need to refer to the decoding key in order to decode the number, thereby reducing the potential for learning effects.

Fourth, in a given round, all participants receive the same decoding key and decode the same sequence of numbers. This ensures all participants experience the task in an identical fashion in every round. Thus, differences in performance cannot be attributable to differences in either different decoding keys or different sequences of numbers in a given round.

2.3.5.3 Task 2: Letter Search Task

In the letter search task, participants count the number of times a search letter, e.g., X, appears in a grid filled with a random array of letters (Sprinkle et al. 2008). For this task, exerting effort entails conducting a thorough scan of the grid. However, as highlighted in Section 2.2, a key concern with the letter search task is participants guessing the correct number of times a search letter appears in a grid rather than conducting a careful count of the search letter. Kachelmeier et al. (2016) find an incorrect count of the search letter often involves participants undercounting the search letter by only one or two instances of the search letter. Thus, if participants have the opportunity to correct their mistakes, an alternative strategy to performing the task is to do a quick scan of the search grid, input an initial count (an educated guess), and then simply adjust the initial count by one or two if the initial count is incorrect. Kachelmeier et al. (2016) find such a strategy may lead to greater performance than thoroughly scanning each grid to do a careful count of the search letter. Most notably, Kachelmeier et al. (2016) find piece-rate pay can motivate participants to adopt this alternative strategy to a greater degree than fixed pay. Thus, differences in performance between these compensation schemes can reflect differences in adopted task strategies and not necessarily effort intensity.

To address this issue, we implemented two design features. First, we used only two possible search letters: X and O. A guessing strategy is more effective when the set of potential search letters is larger (e.g., search letters are all letters A-Z versus only X or O). This occurs because the probability that a cell in the grid contains the search letter decreases as the set of potential search letters increases (assuming a uniform probability distribution), which restricts the variance of the correct count of the search letter. For example, imagine a 5 x 5 grid (25 cells) with a set of potential search letters consisting of all letters A-Z (26 potential search letters), where each cell in the grid is populated with a letter (again, A-Z) using a uniform distribution. With 25 cells each taking on one of 26 potential values (each equally likely), the variance of the correct count of the search letter would be minimal, as it would be very common to have the correct count of the search letter be 0-2 instances. Thus, participants could comfortably guess these frequencies and often be correct. However, if the set of potential search letters is reduced to only two letters, then the variance of the correct count of the search letter is larger. So, the correct count would be less predictable, which reduces the effectiveness of a guessing strategy.

To illustrate, imagine a grid with n cells, and p is the probability that a cell contains the search letter. Assuming independence across cells, the correct count of the search letter follows a binomial distribution with mean np and variance $np(1-p)$. The variance of the correct count of the search letter is maximized when $p = 0.50$. When letters are drawn from the potential set of search letters using a uniform distribution, as in our experiment, this corresponds to a set of only two potential search letters.

Second, we restrict the grid size to one cell. As shown above, the variance of the correct count of the search letter is increasing in the grid size (n), which suggests we should implement a large grid with many cells. At the same time, however, larger grid sizes take longer to scan, which means participants are able to complete fewer grids during a single round. Thus, a smaller grid facilitates a finer unit of performance, and this is maximized if the grid consists of only one cell. Since our version of the letter search task already seeks to discourage guessing by having only two potential search letters, we balance the competing considerations above by implementing grids with only one cell.

Based on these design choices, participants performing our version of the letter search task simply enter the search letter that appears on the screen (either a X or an O).[7] We acknowledge our version of the letter search task (one cell grid with two potential search letters) creates a setting in which participants have a 50% chance to correctly guess the search letter. In fact, one concern is participants will simply press and hold one of the search letters on the computer keyboard, deviating only when they are wrong. However, given the simple nature of the task, there is little to no benefit to guessing the search letter. In addition, recall we assess a

[7] Since there are only two potential search letters, this is functionally identical to typing whether there are zero or one instances of the search letter in the grid.

three-second time penalty for incorrect answers, (along with using only two potential search letters), which should further deter guessing.

A grid size of only one cell is also beneficial because it minimizes the need for participants to actively remember and keep track of instances of the search letter in the grid, thereby minimizing the potential for differences in task-specific skill to explain differences in task performance. In addition, a grid size of only one cell also ensures each grid is equally easy to complete, which increases the likelihood of consistency in task difficulty across rounds.

In our experiment, the program randomly determines whether a X or an O appears on the computer screen. Thus, the exact sequence of letters differs for all participants. However, because of the high volume of task iterations in each round, different sequences of letters in a round across participants is highly unlikely to explain differences in performance.

2.3.5.4 Task 3: Slider Task

In the slider task, participants view a series of sliders with endpoints of 0 and 100, and participants can move the slider box of each slider in increments of 1. The slider box is initially set at 0, and the objective is to use the mouse and drag the slider box to some position along the slider bar (Gill and Prowse 2012, 2015; Chan 2017). In our experiment, participants could attempt up to 48 sliders in each round, and the objective is to move the slider box to the midpoint, i.e., 50. Notably, all sliders in each round are visible on the screen simultaneously so participants do not need to scroll down to complete sliders. While the slider task has been used by accounting research only recently (e.g., Chan 2017), it has been a popular task for experimental economics research (see Gill and Prowse (2015) for a list of studies).

When designing this task, we made six important design choices. First, in each round, we implemented a staggered alignment between adjacent sliders. Gill and Prowse (2015) suggest sliders should not be in close alignment with adjacent sliders to discourage participants from using one correctly positioned slider box as a guide to position the slider box of an adjacent slider. As the intent is for participants to focus on each slider individually, the ability to use adjacent sliders to perform the task introduces an alternative task strategy beyond exerting effort that makes it easier to move a slider box into the correct position.

Second, we changed the alignment of the sliders after every round. Gill and Prowse (2012, 2015) maintain the same alignment for the sliders in every round, arguing this design choice allows them to hold task difficulty constant across rounds. However, they also report significant learning effects, even after multiple practice rounds. Implementing the same alignment in each round may have contributed to the observed learning effects because this design feature allows participants to remember which sliders are "easier" to do (because they are more closely aligned with adjacent sliders). By creating a new alignment for the sliders in each round, we are able to avoid the potential for such effects to emerge.

Third, all sliders in each round are visible on the screen simultaneously instead of being displayed one at a time. Given the two design choices above, one solution for the concerns about slider alignment is to design the task such that only one slider is displayed on the screen at a time, with a new slider appearing only when participants successfully complete a slider. We considered this option, but decided against this approach because of several concerns. For example, if only one slider is displayed on the screen at a time and the position of the slider is fixed for every slider, then learning effects may emerge, similar to those reported by Gill and Prowse (2012, 2015) in which the configuration of sliders was identical in every round. In addition, if only one slider is displayed on the screen at a time and the position of each slider is allowed to vary, then we may introduce another task-specific skill: slider detection.

Fourth, we require participants to manually click and drag the slider box (using the mouse) into the correct position; we disabled the arrow keys on the keyboard *and* prevented participants from clicking on the slider bar to advance the slider box. By forcing participants to move the slider box only by using the mouse, we ensure the act of exerting effort is identical across individuals, thereby further reducing the likelihood of alternative task strategies beyond exerting effort. From our own initial testing, both using the arrow keys and clicking on the slider bar to advance the slider box made the task much easier. By prohibiting these alternative task strategies, we reduce the potential for differences in task strategies to cause differences in performance. Moreover, we reduce the likelihood of learning effects that could emerge from spontaneous discovery and implementation of these alternative task strategies over time. Importantly, prohibiting *both* of these alternative task strategies seems critical. Specifically, Gill and Prowse (2012, 2015) disable the arrow keys on the keyboard, but *allow* participants to click on the slider bar to advance the slider box, and as noted earlier, they observe significant learning effects in their experiment.

Fifth, each correctly positioned slider box counts as one unit of performance. In our experiment, we use this approach to define a unit of performance, as it is the finest unit of performance for the task. Finally, all participants saw the same staggered formation of sliders in a given round. This ensures participants experience the task in an identical fashion over the course of the experiment.

2.4 Results

2.4.1 Overview

In this section, we first test whether we detect incentive effects within each task. Specifically, for each task, we compare the performance between participants earning piece-rate pay to those earning fixed pay. Then, we collapse across contract type conditions (except where noted), and analyze our tasks in terms of the factors discussed in Section 2.2. Three of those factors – fine unit of performance, within-round task experience across participants, and task difficulty across rounds – are effectively addressed by our research design choices; for each task, we chose the finest unit of performance, and within each round, participants experience the task in an identical manner in nearly all of the tasks.[8] Moreover, we expect the difficulty of performing the task to be consistent across rounds because there is a high volume of task iterations in each round, each iteration is randomly generated, and each task iteration is expected to be similar in terms of difficulty. Therefore, we focus our analysis on the four remaining factors: (1) intrinsic motivation; (2) task-specific skill; (3) task strategies beyond exerting effort; and (4) performance trends over time.

[8] As discussed earlier, participants performing the letter search task do not see the exact same sequence of X's and O's in each round. However, we do not expect this to cause inferential difficulties because each letter has a 50 percent chance of appearing and we use a single cell for each letter search grid, which results in a high volume of task iterations in each round.

2.4.2 Incentive Effects

Both economic and psychology theory predict individuals will work harder under a piece-rate contract than under a fixed pay contract. Thus, if our tasks are suitable for detecting variations in participants' effort intensity, then performance will be greater under piece-rate pay than under fixed pay. We test for incentive effects using OLS regressions with robust clustering by participant to address the potential for correlated error terms caused by having eight rounds of performance data for each individual participant (Petersen, 2009).

We examine three models for each task. In Model 1, task performance in each round is the dependent variable, and the independent variable is contract type (0 = fixed pay versus 1 = piece-rate pay). Model 2 is the same as Model 1 except it also controls for participants' practice round performance, which is a (noisy) proxy reflecting participants' baseline performance. Model 3 is the same as Model 2 except it also controls for round fixed effects. When presenting the results of our statistical tests, we report two-tailed p-values unless noted otherwise.[9]

2.4.2.1 Decode Task

Table 2, Panel A, reports the descriptive statistics for participants' task performance within each round and across all eight performance rounds for the fixed pay and the piece-rate pay conditions. Mean (standard deviation) performance across all eight performance rounds is

[9] We also test for incentive effects using random-effects regressions, as this approach also addresses the potential for correlated error terms and understated standard errors caused by having eight rounds of individual performance data nested within each individual participant (Klein and Kozlowski 2000; Luke 2004). We conduct two sets of regressions. In the first set, task performance in each round is the dependent variable, and the independent variables are contract type (0 = fixed pay versus 1 = piece-rate pay), round (eight rounds), and the interaction between contract type and round. The second set of regressions also controls for participants' practice round performance. The results of these random-effects regressions are inferentially similar to those of our main OLS regressions.

24.92 (4.66) in the fixed pay condition and 26.63 (4.16) in the piece-rate pay condition. As reported in Table 2, Panel B, the regression results indicate participants in the piece-rate pay condition marginally outperformed those in the fixed pay condition (all one-tailed p-values $\leq$ 0.10). Thus, we find marginal evidence of incentive effects using the decode task.

2.4.2.2 Letter Search Task

Table 3, Panel A, reports the descriptive statistics for participants' task performance within each round and across all eight performance rounds for the fixed pay and the piece-rate pay conditions. Mean (standard deviation) performance across all eight performance rounds is 162.71 (16.95) in the fixed pay condition and 134.64 (12.23) in the piece-rate pay condition. Surprisingly, regression results reported in Table 3, Panel B, indicate participants in the fixed pay condition *outperformed* those in the piece-rate pay condition (all p-values < 0.01).[10] Thus, we do not detect incentive effects using the letter search task. In fact, we find the opposite result.

Interestingly, this performance difference occurred even though participants in the fixed pay condition committed significantly more errors than participants in the piece-rate pay condition (7.64 vs. 5.32, respectively; $t = 3.24$, $p < .01$). This difference in error rates implies participants in the fixed pay condition had seven fewer seconds per round (out of a possible 120 seconds) to perform the task. So, participants in the fixed pay condition completed more units even though they had less time per round to perform the task.

[10] We report two-tailed p-values here because results are in the opposite direction of our expectation.

Table 2: Decode Task Results[a]

Panel A: Mean (Standard Deviation) Performance[b]

	Practice Round	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7	Round 8	Average
Piece-rate	22.32 (5.91)	27.35 (5.00)	27.12 (5.20)	26.68 (4.26)	27.79 (5.26)	26.06 (4.79)	25.32 (5.22)	27.38 (4.81)	25.35 (4.98)	26.63 (4.16)
Fixed	20.97 (5.13)	25.15 (6.02)	26.12 (5.53)	24.64 (6.09)	26.61 (4.70)	24.39 (6.24)	24.12 (5.32)	25.24 (6.36)	23.09 (5.00)	24.92 (4.66)

Panel B: OLS Regressions[c]

	Model 1	Model 2	Model 3
Intercept	25.15 (t = 30.92, p < 0.01)	12.26 (t = 7.81, p < 0.01)	12.49 (t = 10.71, p < 0.01)
Contract Type (0 = fixed pay, 1 = piece-rate pay)	1.71 (t = 1.60, one-tailed p = 0.06)	0.89 (t = 1.31, one-tailed p = 0.10)	0.89 (t = 1.30, one-tailed p = 0.10)
Practice Performance		0.60 (t = 8.82, p < 0.01)	0.60 (t = 8.77, p < 0.01)
Round Fixed Effects	No	No	Yes
R^2	2.5%	39.7%	42.7%

[a] Sixty-seven participants completed the decode task, in which the decoding key is presented in alphabetical order. Participants performed a two-minute practice round, followed by eight two-minute main performance rounds. Participants in the piece-rate pay condition earned \$0.35 for each correctly decoded number in each round. Participants in the fixed pay condition earn \$9.35 in each round, regardless of performance. We randomly selected one of the main performance rounds for payment. Thirty-four participants worked under the piece-rate pay contract, and thirty-three participants worked under the fixed pay contract. We used average pay under the piece-rate pay contract as the compensation amount for the fixed pay contract, ensuring average pay is equivalent between contract type conditions.

[b] In each round, performance is the amount of correctly decoded numbers in the round. The “Average” column reports the average performance per round across the eight main performance rounds, and does not include practice round performance.

[c] The analysis includes 536 observations (67 participants x 8 rounds per participant). All three models include robust clustering by individual to protect against correlated error terms. In Model 1, task performance in each round is the dependent variable, and the independent variable is contract type (0 = fixed pay vs. 1 = piece-rate pay). Model 2 is the same as Model 1 except it also controls for participants’ practice round performance. Model 3 is the same as Model 2 except it controls for round fixed effects (i.e., a dummy variable). All p-values are reported two-tailed unless noted otherwise. All three models report the unstandardized regression coefficients, t-statistics, and p-values.

It is possible participants in the piece-rate pay condition perceived our three-second penalty for incorrect entries as too punitive. On average, participants completed one unit of performance in less than one second, which means the three-second penalty is quite costly, especially in the piece-rate pay condition. So, avoiding mistakes was likely quite salient to participants in the piece-rate pay condition, and they may have reduced the speed at which they were performing the task to ensure they did not make costly mistakes, which lowered their performance.[11]

2.4.2.3 Slider Task

Table 4, Panel A, reports the descriptive statistics for participants' task performance within each round and across all eight performance rounds for the fixed pay and the piece-rate pay conditions. The descriptive statistics and the analyses reported below exclude data from one participant.[12] Mean (standard deviation) performance across all eight performance rounds is 31.27 (7.95) in the fixed pay condition and 36.52 (4.68) in the piece-rate pay condition. As reported in Table 4, Panel B, the regression results indicate participants in the piece-rate pay condition outperformed those in the fixed pay condition (all one-tailed p-values < 0.01). Thus, we detect incentive effects using the slider task.

---

[11] We also examine participants' responses to an open-ended item from the post-experimental questionnaire that asks participants to describe strategies they used to perform well on the task. We code the responses as a one if participants explicitly mentioned avoiding the time penalty in their strategy description, and zero otherwise. Although we find participants in the piece-rate pay condition mentioned avoiding the penalty more frequently than do participants in the fixed pay condition, we do not find this difference is statistically significant at conventional levels (9.38 percent versus 0.00 percent, Fisher's exact test p = 0.24)

[12] One participant in the piece-rate pay condition completed an average of only 0.25 sliders per round, which is 4.40 standard deviations below the mean. Including the data from this participant would have lowered average pay, and thus the fixed pay amount, by $0.40. So, we chose to exclude this participant's data before collecting data in the fixed pay condition.

Table 3: Letter Search Task Results[a]

Panel A: Mean (Standard Deviation) Performance[b]

	Practice Round	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7	Round 8	Average
Piece-rate	124.63 (25.51)	140.56 (12.20)	136.13 (13.54)	136.47 (12.77)	132.47 (11.32)	133.63 (13.27)	134.13 (14.80)	131.41 (27.53)	132.38 (15.44)	134.64 (12.23)
Fixed	153.67 (48.07)	173.59 (19.97)	163.67 (18.98)	161.19 (17.95)	160.04 (23.21)	159.11 (23.87)	161.26 (20.14)	162.78 (20.46)	160.07 (24.66)	162.71 (16.95)

Panel B: OLS Regressions[c]

	Model 1	Model 2	Model 3
Intercept	162.71 (t = 50.34, p < 0.01)	144.95 (t = 9.03, p < 0.01)	12.49 (t = 10.71, p < 0.01)
Contract Type (0 = fixed pay, 1 = piece-rate pay)	-28.07 (t = -7.23, p < 0.01)	-24.71 (t = -5.03, p < 0.01)	-24.71 (t = -4.99, p < 0.01)
Practice Performance		0.12 (t = 1.21, p = 0.23)	0.12 (t = 1.20, p < 0.23)
Round Fixed Effects	No	No	Yes
R^2	36.2%	39.6%	41.6%

[a] Fifty-nine participants completed the letter search task. Participants performed a two-minute practice round, followed by eight two-minute main performance rounds. Participants in the piece-rate pay condition earned \$0.10 for each correctly decoded number in each round. Participants in the fixed pay condition earn \$13.50 in each round, regardless of performance. We randomly selected one of the main performance rounds for payment. Thirty-two participants worked under the piece-rate pay contract, and twenty-seven participants worked under the fixed pay contract. We used average pay under the piece-rate pay contract as the compensation amount for the fixed pay contract, ensuring average pay is equivalent between contract type conditions.

[b] In each round, performance is the amount of typed letters. The "Average" column reports the average performance per round across the eight main performance rounds, and does not include practice round performance.

[c] The analysis includes 472 observations (59 participants x 8 rounds per participant). All three models include robust clustering by individual to protect against correlated error terms. In Model 1, task performance in each round is the dependent variable, and the independent variable is contract type (0 = fixed pay vs. 1 = piece-rate pay). Model 2 is the same as Model 1 except it also controls for participants' practice round performance. Model 3 is the same as Model 2 except it controls for round fixed effects (i.e., a dummy variable). All p-values are reported two-tailed because results are in the opposite direction of our expectation. All three models report the unstandardized regression coefficients, t-statistics, and p-values.

Table 4: Slider Task Results[a]

Panel A: Mean (Standard Deviation) Performance[b]

	Practice Round	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7	Round 8	Average
Piece-rate	30.53 (8.52)	35.27 (4.14)	35.37 (4.77)	35.53 (5.54)	37.37 (4.87)	37.10 (5.55)	37.33 (5.39)	36.97 (5.79)	37.23 (5.37)	36.52 (4.68)
Fixed	30.16 (7.76)	33.56 (8.09)	32.00 (7.82)	32.38 (9.07)	32.22 (9.68)	29.84 (10.90)	30.38 (11.34)	29.03 (11.75)	30.72 (12.45)	31.27 (7.95)

Panel B: OLS Regressions[c]

	Model 1	Model 2	Model 3
Intercept	31.27 (t = 22.39, p < 0.01)	19.03 (t = 4.92, p < 0.01)	12.49 (t = 10.71, p < 0.01)
Contract Type (0 = fixed pay, 1 = piece-rate pay)	5.22 (t = 3.22, one-tailed p < 0.01)	5.10 (t = 3.56, one-tailed p < 0.01)	5.10 (t = 3.53, one-tailed p < 0.01)
Practice Performance		0.41 (t = 3.24, p < 0.01)	0.41 (t = 3.22, p < 0.01)
Round Fixed Effects	No	No	Yes
R^2	9.4%	23.7%	24.1%

[a] Sixty-three participants completed the slider task. Participants performed a two-minute practice round, followed by eight two-minute main performance rounds. Participants in the piece-rate pay condition earned \$0.30 for each correctly decoded number in each round. Participants in the fixed pay condition earn \$11 in each round, regardless of performance. We randomly selected one of the main performance rounds for payment. Thirty-one participants worked under the piece-rate pay contract, and thirty-two participants worked under the fixed pay contract. We used average pay under the piece-rate pay contract as the compensation amount for the fixed pay contract, ensuring average pay is equivalent between contract type conditions. We excluded data from one participant who completed an average of only 0.25 sliders per round, which is 4.40 standard deviations below the mean. Thus, our analysis includes data from the remaining sixty-two participants.

[b] In each round, performance is the amount of correctly positioned sliders. The "Average" column reports the average performance per round across the eight main performance rounds, and does not include practice round performance.

[c] The analysis includes 496 observations (62 participants x 8 rounds per participant). All three models include robust clustering by individual to protect against correlated error terms. In Model 1, task performance in each round is the dependent variable, and the independent variable is contract type (0 = fixed pay vs. 1 = piece-rate pay). Model 2 is the same as Model 1 except it also controls for participants' practice round performance. Model 3 is the same as Model 2 except it controls for round fixed effects (i.e., a dummy variable). All p-values are reported two-tailed unless noted otherwise. All three models report the unstandardized regression coefficients, t-statistics, and p-values.

2.4.3 Analysis of Factors

We now analyze the tasks in terms of the factors discussed in Section 2.2. As noted earlier, the analyses in this subsection focus on four factors: (1) intrinsic motivation; (2) task-specific skill; (3) task strategies beyond exerting effort; and (4) performance trends over time. Although we detect incentive effects for only two of the tasks (the decode task and the slider task), we report analyses regarding the four factors for all tasks.

2.4.3.1 Intrinsic Motivation

As previously discussed, participants with greater intrinsic motivation are less likely to be effort-averse, which creates an upper bound issue in terms of effort and performance. Participants may be intrinsically motivated to exert effort on some real-effort tasks because those tasks are interesting in nature and/or performing those tasks well is personally important. To test whether intrinsic motivation in each task is sufficiently low, we analyze participants' responses to five items from the post-experimental questionnaire: (1) The task is attractive; (2) The task is fun; (3) The task is exciting; (4) The task is interesting; and (5) Doing well on the task is important to me. For each item, participants rated their level of agreement using a 7-point scale with endpoints "Strongly Disagree" (-3) and "Strongly Agree" (+3), and a midpoint of "Neither Agree nor Disagree" (0).

The first four items above test whether participants found that the task was interesting. To assess whether responses to these items reflect a single underlying construct, we conduct a factor analysis on these four items, separately for each task. All factor loadings for all tasks are greater than 0.74 and the eigenvalues (variance explained) are greater than 2.71 (68 percent). In addition, the Cronbach's alpha is greater than 0.89 for each task. Thus, we average the responses for the four items to create a single factor, which we label *Interest*.

Table 5 reports descriptive statistics for *Interest* for each task. Mean *Interest* across tasks varies from a low of -0.34 for the letter search task to a high of 0.33 for the decode task. For all three tasks, *Interest* is significantly higher than the minimum possible value of -3 (all t-statistics > 13.53 and all p-values < 0.01). *Interest* is also significantly lower than the maximum possible value of +3 for all three tasks (all t-statistics > 13.66 and all p-values < 0.01). Further, *Interest* is either below or statistically similar to the neutral value of 0 for the letter search task and the slider task, but it is above the neutral value of 0 for the decode task (decode task: $t = 1.68$, $p = 0.10$; letter search task: $t = 1.72$, $p = 0.09$; slider task: $t = 1.29$, $p = 0.20$).

The fifth item above captures the extent to which participants feel performing well on the task is personally important to them (*Important*). Table 5 reports descriptive statistics for *Important* for each task. Mean *Important* across tasks varies from a low of 0.56 for the letter search task to a high of 0.75 for the decode task. For all three tasks, *Important* is significantly higher than the neutral value of 0 (all t-statistics > 2.38 and all p-values < 0.03), but significantly lower than the maximum possible value of +3 (all t-statistics > 9.95 and all p-values < 0.01).[13]

These results suggest our design of all three tasks effectively muted participants' intrinsic motivation to perform those tasks, which provides greater scope for detecting incentive effects, should they exist.[14]

---

[13] We also explore the correlation between *Interest* and *Important*. For all three tasks, the correlation is positive and statistically significant (decode task: $r = 0.47$, $p < 0.01$; letter search task: $r = 0.35$, $p < 0.01$; slider task: $r = 0.48$, $p < 0.01$).

[14] Theory predicts intrinsic motivation will positively affect performance. We find *Interest* is positively correlated with mean performance for the decode task ($r = 0.25$, $p = 0.04$) and the slider task ($r = 0.33$, $p < 0.01$), but not for the letter search task ($r = .04$, $p = 0.79$). We also find *Important* is positively correlated with mean performance for the slider task ($r = 0.38$, $p < 0.01$), but not for the decode task ($r = 0.16$, $p = 0.19$) or the letter search task ($r = .04$, $p = 0.77$).

Table 5: Factors Discussed in Section 2.2

Task	Intrinsic Motivation: *Interest*[a]	Intrinsic Motivation: *Important*[b]	Task-Specific Skill: *Skill Gap*[c]	Task Strategies Beyond Exerting Effort[d]	Fine Unit of Performance[e]	Within-Round Task Experience Across Participants[e]	Performance Trends Over Time: *Fatigue*[f]	Performance Trends Over Time: *Learning Effects*[g]	Task Difficulty Across Rounds[e]
Decode	0.33 (1.60)	0.82 (1.79)	0.06 (1.43)	13.43%	N/A	N/A	0.74 (1.64)	$\beta = -0.22$ $p = 0.01$	N/A
Letter Search	-0.34 (1.51)	0.74 (1.77)	-0.82 (1.20)	0.00%	N/A	N/A	1.34 (1.64)	$\beta = -1.03$ $p = 0.01$	N/A
Slider	-0.24 (1.45)	0.56 (1.80)	-1.01 (1.36)	8.06%	N/A	N/A	0.47 (2.15)	$\beta = 0.32$ $p < 0.01$	N/A

[a] There are two antecedents to intrinsic motivation: task interest and task importance. We report the mean (standard deviation) of *Interest,* which is the average of participants' responses to four items from the post-experimental questionnaire: (1) The task is attractive; (2) The task is exciting; (3) The task is interesting; and (4) The task is fun.

[b] We report the mean (standard deviation) of *Important, which* is the average of participants' responses to the post-experimental questionnaire item: Doing well on the task is important to me. For each item, participants rate their level of agreement using a 7-point scale with endpoints of "Strongly Disagree" (-3) and "Strongly Agree" (+3), and a midpoint of "Neither Agree nor Disagree" (0).

[c] We report the mean (standard deviation) of *Skill Gap, which* is the average of participants' responses to three items from the post-experimental questionnaire: (1) The task is complex; (2) The task is mentally demanding; and (3) The task requires thought and problem solving. For each item, participant rate their level of agreement using a 7-pt scale with end points of "Strongly Disagree" (-3) and "Strongly Agree" (+3), and a midpoint of "Neither Agree nor Disagree" (0).

[d] In the post-experimental questionnaire, participants respond to the following prompt: Please describe any strategy you used to help you perform well on the task. We code each participant's response as zero if it described performing the task in the manner in which we intended, and as one if it described any other strategy. The reported percentage is the percentage of participants who describe an alternative strategy.

[e] For fine unit of performance, within-round task experience across participants, and task difficulty across rounds, we believe we have effectively addressed these through design choices. Specifically, for each task, we chose the finest unit of performance, and within each round, participants experience the task in an identical manner in nearly all of the tasks. Moreover, we expect the difficulty of performing the task to be consistent across rounds because there is a high volume of task iterations in each round, each iteration is randomly generated, and each task iteration is expected to be similar in terms of difficulty.

[f] We report the mean (standard deviation) of *Fatigue,* which is participants' response to the following item from the post-experimental questionnaire: I felt tired at the end of the eighth round due to my work on the task. Participants rate their level of agreement using a 7-point scale with endpoints of "Strongly Disagree" (-3) and "Strongly Agree" (+3), and a midpoint of "Neither Agree nor Disagree" (0). We only include responses from participants in the piece-rate pay condition as this contract type offers participants an explicit incentive to exert effort in order to maximize their earnings.

[g] To measure *Learning Effects,* we conduct OLS regressions with robust clustering by participant for the piece-rate pay conditions with performance from the main performance rounds as the dependent variable and round (1-8) as the independent variable. We report the coefficient for round and its associated p-value.

2.4.3.2 Task-Specific Skill

Participants' task-specific skill can have a substantive impact on task performance. As discussed in Section 2.2, a key aspect of task-specific skill is ensuring all participants have (or at least perceive they have) the requisite skill to perform the task. We expect participants who perceive they lack the requisite skills will perceive the task as more difficult. To test whether participants felt they had sufficient skill to complete the tasks, we analyze participants' responses to three items from the post-experimental questionnaire: (1) The task is complex; (2) The task is mentally demanding; and (3) The task requires thought and problem-solving. For each item, participants rated their level of agreement using a 7-point scale with endpoints "Strongly Disagree" (-3) and "Strongly Agree" (+3), and a midpoint of "Neither Agree nor Disagree" (0).

To assess whether responses to these three items reflect a single underlying construct, we conduct a factor analysis on these three items, separately for each task. All factor loadings are greater than 0.44 and the eigenvalues (variance explained) are greater than 1.04 (35 percent). In addition, the Cronbach's Alpha for these three items is greater than 0.65 for each task. Thus, we find these items represent a single factor, which we label *Skills Gap.* We calculate participants' *Skills Gap* by averaging their responses to the three items.

Table 5 reports the descriptive statistics for *Skills Gap* for each of the three tasks. Mean *Skills Gap* across tasks varies from a low of -1.01 for the slider task to a high of 0.06 for the decode task. For all three tasks, *Skills Gap* is significantly higher than the minimum possible value of -3 (all t-statistics > 11.52 and all p-values < 0.01). *Skills Gap* is also significantly lower than the maximum possible value of +3 (all t-statistics > 16.83 and all p-values < 0.01). Further, except for the decode task ($t = 0.34$, $p = 0.73$), *Skills Gap* is less than the neutral value of 0 (letter search task: $t = 5.28$, $p < 0.01$; slider task: $t = 5.81$, $p < 0.01$). These results suggest participants felt they had sufficient skill to perform each of the tasks, but more so for the letter search task and the slider task.[15]

2.4.3.3 Task Strategies Beyond Exerting Effort

When using an effort-intensive task, it is important to ensure participants who choose to exert effort do so only in the matter the researchers intended. In the post-experimental questionnaire, participants responded to the following open-ended prompt: “Please describe any strategy you used to help you perform well on the task.” We coded each participant’s response as a zero if the response described performing the task in the manner in which we intended, and a one if the response described an alternative task strategy.

Table 5 presents the percentage of participants describing an alternative task strategy, separately for each task. For the decode task, 13.43 percent of participants described an alternative strategy of trying to memorize some portion of the decoding key to improve future performance, even though we tried to discourage this strategy by changing the decoding key at the end of every round and using two-minute performance rounds. For the letter search task, no participants described any alternative strategies. For the slider task, 8.06 percent of participants described an alternative strategy of using the relative position of other sliders to help complete the slider they were currently working on, even though we tried to discourage this strategy by implementing a staggered formation of adjacent sliders. Overall, it appears the design of our tasks was effective in limiting alternative task strategies beyond exerting effort, as the percentages reported above range from a low of 0 percent to a high of 13.43 percent. Moreover,

---

[15] Theory predicts a lack of requisite skills will lead to lower performance. Interestingly, we find *Skills Gap* is not correlated with mean performance for the decode task ($r = -0.03$, $p = 0.79$) or the letter search task ($r = -0.04$, $p = 0.76$), but is positively correlated with mean performance for the slider task ($r = 0.24$, $p = 0.06$).

it appears the designs of the letter search and slider tasks were more effective in limiting alternative task strategies than the decode task.

2.4.3.4 Performance Trends Over Time

When studies involve multiple rounds, participants gain substantial practice, and thus, greater comfort and familiarity with the task, through repeated exposure. As such, participants' performance may improve over time simply due to practice (i.e., learning effects) and not due to increased effort. We test for learning effects by examining performance over time within the piece-rate pay condition, as this contract type offers participants an explicit incentive to exert effort in order to maximize their earnings. Specifically, we conduct OLS regressions with robust clustering by participant (untabulated) for the piece-rate pay condition with performance from the main performance rounds as the dependent variable, and round (1-8) as the independent variable. We find evidence consistent with learning effects for only the slider task, where average performance increases by 1.96 units from round 1 to round 8 ($\beta = 0.32$, $p < 0.01$). For the other two tasks, we find a downward trend in performance over time (decode task: $\beta = -0.22$, $p = 0.01$; letter search task: $\beta = -1.03$, $p = 0.01$). Thus, unlike the design of the slider task, it appears the design of the other two tasks was effective in limiting learning effects.

The downward trend in performance over time for two of our tasks may reflect fatigue. Thus, we analyze participants' responses to the following item from the post-experimental questionnaire: I felt tired at the end of the eighth round due to my work on the task. Participants rate their level of agreement with this item using a 7-point scale with endpoints "Strongly Disagree" (-3) and "Strongly Agree" (+3), and a midpoint of "Neither Agree nor Disagree" (0). Our measure, *Fatigue*, is participants' responses to the item above.

Table 5 reports descriptive statistics for *Fatigue* for only those in the piece-rate pay condition as this contract type offers participants an explicit incentive to exert effort in order to maximize their earnings. Mean *Fatigue* is 0.47 for the decode task, 0.74 for the slider task, and 1.34 for the letter search task. For all tasks, *Fatigue* is significantly greater than the minimum possible value of -3 (all t-statistics > 8.85 and all p-values < 0.01). *Fatigue* is also significantly below the maximum possible value of +3 (all t-statistics > 5.72 and all p-values < 0.01). Further, *Fatigue* is not different from the neutral value of 0 for the slider task ($t = 1.19$, $p = 0.24$), but is greater than the neutral value of 0 for the decode task ($t = 2.61$, $p = 0.01$) and the letter search task ($t = 4.64$, $p < 0.01$).[16]

2.5 Conclusion

Researchers commonly use real-effort tasks to investigate the effects of incentive pay on effort intensity in a laboratory experiment. We identify seven factors researchers should consider when designing an effort-intensive task: (1) intrinsic motivation; (2) task-specific skill; (3) task strategies beyond exerting effort; (4) fine unit of performance; (5) within-round task experience across participants; (6) performance trends over time; and (7) task difficulty across rounds. With these factors in mind, we design an experiment to test for incentive effects on effort using several tasks used in prior accounting research. Contrary to our expectation of observing incentive effects for all of tasks in our experiment, we only detect incentive effects for the decode task and the slider task, but not for the letter search task. Regarding the seven factors, the global and task-

[16] We would expect greater fatigue to negatively affect performance. For all tasks, we find *Fatigue* is not correlated with mean performance (all p-values > 0.40).

specific features we included in our experiment were generally effective in limiting the effects of those factors, though with varying levels of success, depending on the factor and the task in question. Collectively, the strongest evidence for incentive effects and for limiting the effects of other factors comes from our version of the slider task.

We believe our study contributes to research by helping researchers who are considering using effort-intensive tasks. Our list of factors will help researchers design tasks that allow them to conduct a more effective test of theory due to a stronger link between effort intensity and performance. We also highlight the difficulty of using real-effort tasks to detect incentive effects in laboratory experiments, and our analysis of design factors offer some insights into why this might be. Finally, we provide evidence the slider task may be the most effective task for researchers to use when testing the effect of incentives on effort intensity.

3.0 Do Vesting Requirements Increase the Incentive Effects of Stock Compensation for Rank-and-File Employees?

3.1 Introduction

Firms commonly reward rank-and-file employees with stock-based compensation. The National Center for Employee Ownership (NCEO) estimates 32 million Americans (approximately one-fourth of working adults) own equity in their company (NCEO 2014). In addition, survey responses indicate fifty-one percent of rank-and-file employees at publicly-traded companies are now incentivized with stock compensation, which is normally granted with vesting requirements (WorldAtWork 2014). In this same survey, ninety-three percent of respondents indicated that a primary purpose of offering stock-based compensation is to align employee incentives with organizational goals, causing employees to provide more effort.[17] Despite the prevalence of these stock compensation arrangements with vesting requirements, and survey results indicating that firms use them to motivate employee effort, little is known about how they actually affect rank-and-file employees' effort. My study addresses this gap in the literature by examining whether stock compensation with vesting requirements motivates employees more effectively than other compensation arrangements, which would help explain why such stock compensation arrangements are so common for rank-and-file employees.

[17] The focus of the current study is on the effect of stock compensation on employee effort. Research has investigated other potential benefits firms receive by offering stock-based compensation such as sorting/attracting employees (Hales, Wang, and Williamson 2015), retaining employees (Ittner, Lambert, and Larcker 2003; Oyer and Schaefer 2005), and conserving cash when facing financing constraints (Core and Guay 2001).

Stock with vesting requirements might be a particularly effective incentive because of how it is framed in compensation contracts. Research on framing has defined a bonus-framed contract as one in which the employee receives a bonus for meeting a specified target and a penalty-framed contract as one in which the employee incurs a penalty for failing to meet a specified target. I predict that employees who receive stock with vesting requirements (restricted stock) will likely perceive this as a penalty-framed contract because although they initially own the stock, they will lose it if they do not meet the vesting requirements. On the other hand, I predict employees who receive stock without vesting requirements (unrestricted stock) will likely perceive this as a bonus-framed contract because they do not initially own the stock and have the opportunity to earn it. Although neoclassical economic theory assumes that the framing of equivalent contracts in terms of gains or losses has no impact on individuals' behavior, research suggests employees provide more effort under penalty-framed contracts than under economically equivalent bonus-framed contracts (Hannan, Hoffman, and Moser 2005; Church, Libby, and Zhang 2008). Greater effort under penalty contracts has been attributed to loss aversion, i.e., individuals react more negatively to losses than they react favorably to equivalent gains (Kahneman and Tversky 1984).

Studies examining the effect of bonus or penalty framing on effort have used cash as the form of compensation for the bonus or penalty amount. Beyond any basic effect of a bonus versus penalty frame on effort, as identified in prior literature and described above, I propose the motivating effect of a penalty frame on effort will differ between cash and stock compensation due to the endowment effect. The endowment effect describes the finding that individuals value an item more when it is owned, i.e., when it is part of their endowment, than when it is not owned (Thaler 1980). Employees working under a penalty-framed contract own the penalty portion of their compensation (i.e., it is part of their endowment) and should therefore assign a higher value to it due to the endowment effect. In contrast, employees working under a bonus-framed contract do not yet own the bonus portion of their compensation, so its value is unaffected by the endowment effect. Therefore, consistent with the endowment effect, I expect employees will place a higher value on their compensation under a penalty-framed contract than a bonus-framed contract. However, as described below, this increase in value will occur when the penalty-framed compensation is stock, but not cash.

Kahneman, Knetsch, and Thaler (1990) identify a critical exception to the endowment effect, such that it does not affect valuations of cash because cash is difficult to mentally revalue. For example, individuals do not re-value $10 of cash as being worth more than $10 whether they own the $10 or not. Therefore, I expect the endowment effect to differentially affect how employees value stock and cash in penalty-framed contracts. Specifically, the endowment effect will cause employees to increase the value they assign to stock when it is the penalty amount in a penalty-framed contract (hereafter referred to as a penalty-framed stock contract), but to have no effect on the value of cash when it is the penalty amount in a penalty-framed contract (hereafter referred to as a penalty-framed cash contract). The value of the stock in a penalty-framed contract is also expected to be higher than the value of either cash or stock in a bonus-framed contract because employees do not own those potential bonus amounts. Because employees work harder for compensation they value more highly, I predict effort will be highest under the penalty-framed stock contract, which represents stock with vesting requirements. This result can help explain why restricted stock with vesting requirements is common in rank-and-file compensation contracts.

Before conducting my main experiment, I collect survey data confirming my expectation that individuals perceive restricted stock compensation arrangements with vesting requirements as penalty contracts and unrestricted stock compensation arrangements without vesting requirements as bonus contracts. I then conduct a 2 x 2 between-participants experiment in which I manipulate the compensation form (cash versus stock) and the contract frame (bonus versus penalty). As predicted, I find effort is highest under the penalty-framed stock contract, reflecting the fact that the effect of a penalty frame on effort is greater when the penalty amount is in the form of stock than in the form of cash. Moreover, consistent with the underlying theory for the hypothesized endowment effect, participants working under the penalty-framed stock contract value their stock compensation higher than participants working under the bonus-framed stock contract, and this valuation mediates the effect of contract frame on effort that I find for the stock contracts.

My study makes three important contributions. First, I demonstrate unrestricted stock is perceived as a bonus whereas restricted stock with vesting requirements is perceived as a penalty. Second, I conceptually link the framing and endowment effect literatures by introducing the idea that in penalty-framed contracts, employees are endowed with the penalty portion of the contract and thus feel they own it. This ownership causes individuals to value the compensation more highly when it is in the form of stock, but not when it is in the form of cash because cash is difficult to mentally re-value. Third, I demonstrate that stock contracts with features comparable to vesting requirements lead to more effort than the other types of economically equivalent contracts. This helps explain why restricted stock is so prevalent in rank-and-file compensation contracts.

The remainder of this paper is organized as follows. Section 3.2 describes my theory and develops my hypothesis. Section 3.3 describes my method and Section 3.4 details my results. I offer concluding remarks in Section 3.5.

3.2 Theory and Hypothesis

3.2.1 Stock-Based Compensation

Firms are increasingly using stock-based compensation to incentivize rank-and-file employees to provide effort. This is surprising given that economic theory predicts stock-based compensation cannot effectively incentivize rank-and-file employees for several reasons, including employee risk aversion (Goetzmann and Kumar 2008), elevated levels of free-riding (Hall and Murphy 2003), and the inability of individual rank-and-file employees to affect stock price (Merchant and Van der Stede 2003). Indeed, Hall and Murphy (2003, p. 58) conclude, "it seems obvious that cash-based incentive plans based on objective or subjective performance measures can provide stronger and more efficient pay-performance incentives."

However, despite this economic reasoning, the available archival evidence suggests the incentive effects of stock-based compensation are at least as strong as those of cash (Blasi, Kruse, and Sesil 2001; Hochberg and Lindsey 2010; Kim and Ouimet 2014). Blasi et al. (2003, p. 98) review findings from over thirty studies and conclude that although broadly granting stock-based compensation to rank-and-file employees does not automatically improve firm performance, it appears "there may be benefits – and are unlikely to be adverse consequences - from the expansion of employee equity stakes in companies." In this study, I follow up on Blasi et al.'s (2003) call to identify when stock is a particularly effective incentive, such that it can motivate rank-and-file employees to provide more effort than other economically equivalent incentives.

Two common forms of stock-based compensation that firms offer to rank-and-file employees are unrestricted stock and restricted stock with vesting requirements. Unrestricted stock is stock that immediately transfers to an employee after being earned with no restrictions on ownership, whereas restricted stock is stock that is granted to an employee, but that only vests (is retained and may be sold) if the employee meets specific requirements.[18] The failure to meet these vesting requirements causes the employee to forfeit the restricted stock. I investigate whether the use of vesting requirements in restricted stock contracts is an important factor in explaining when stock is a particularly effective incentive.

Unrestricted stock and restricted stock have features that are functionally equivalent to bonus-framed and penalty-framed contracts described in the contract-framing literature (Hannan et al. 2005). I argue that employees who receive stock without vesting requirements (unrestricted stock) likely perceive this as a bonus arrangement because they do not initially own the stock and have the opportunity to earn it. In contrast, employees who receive stock with vesting requirements (restricted stock) likely perceive this as a penalty arrangement because they initially own the stock, but will have to forfeit it if they do not meet the vesting requirements.

[18] The most common type of vesting requirement is time vesting (e.g., the stock vests if the employee still works for the company three years later). This type of vesting requirement has an element of performance, as employees must perform at a sufficient level to have continued employment. However, a more recent trend is to tie vesting requirements explicitly to performance outcomes. For example, a sales person may need to meet a sales target for their stock to vest (Eppert 2017). Indeed, some firms now call this type of incentive arrangement "performance shares" instead of restricted stock to explicitly indicate the difference. In a recent survey, respondents indicated fifteen percent of rank-and-file employees at publicly-traded companies are now incentivized by performance shares/units (WorldAtWork 2014).

3.2.2 Contract Frame: Bonus Versus Penalty

I use the term contract frame to refer to whether a contract is framed as a bonus or a penalty. A bonus-framed contract is one in which the employee receives a bonus for meeting a specified target whereas a penalty-framed contract is one in which the employee incurs a penalty for failing to meet a specified target. For example, in a bonus-framed contract, an employee may receive a base payment of $7 and earn a bonus of $5 for meeting a specified target. In an economically equivalent penalty-framed contract, the employee would receive a base payment of $12 and incur a penalty of $5 for failing to meet the same specified target. Despite the different contract frames (bonus versus penalty), these two contracts are economically equivalent because under both contracts employees who meet the specified target will receive $12 and employees who fail to meet the specified target will receive $7. Neoclassical economic theory predicts that employee behavior will not differ between two such economically equivalent contracts despite their different contract frames (Demski and Feltham 1978).

Despite this economic prediction, Luft (1994) finds participants strongly prefer a bonus-framed contract to an economically equivalent penalty-framed contract. However, her study cannot speak to whether contract frame affects effort because her participants perform a memory task for which ability (knowledge), and not effort, is the main determinant of task performance. In contrast, Hannan et al. (2005) extend Luft's (1994) study by using an effort-sensitive task in their experiment, and find that participants work harder under a penalty-framed contract than an economically equivalent bonus-framed contract. Similarly, Lazear (1991) argues penalty-framed contracts could lead to greater effort than bonus-framed contracts, despite economic equivalence, because of loss aversion (Kahneman and Tversky 1984). Loss aversion is an important part of prospect theory (Kahneman and Tversky 1979), in which individuals evaluate financial outcomes as gains or losses relative to some reference point. Even when bonus-framed and penalty-framed contracts are economically equivalent, individuals tend to frame the bonus portion of a bonus-framed contract as a potential gain, *incremental to the reference point*, while they tend to frame the penalty portion of a penalty-framed contract as *part of the (now higher) reference point.* Therefore, under the penalty-framed contract, individuals will view having to pay the penalty as a loss. According to prospect theory, losses loom larger than gains, i.e., individuals react more negatively to losses than they react favorably to equivalent gains. Therefore, employees likely perceive that more value is lost by incurring a penalty than is gained by earning an equivalent bonus. Given this theory and considerable supporting evidence, I expect penalty-framed contracts to lead to greater effort than economically equivalent bonus-framed contracts.

Finding greater effort under penalty-framed contracts than bonus-framed contracts would replicate Hannan et al. (2005), who, like other researchers in the area, used cash as the form of compensation for the bonus and penalty portions of the contracts. However, as explained in the next subsection, I predict the motivating effect of a penalty frame (as opposed to a bonus frame) will differ between cash and stock compensation due to the endowment effect. Thus, the relative incentive effects of cash versus stock may differ depending on whether they are framed as a bonus or as a penalty.

3.2.3 Endowment Effect

The endowment effect describes the finding that individuals value an item higher if it is owned, or part of their "endowment," than if it is not owned (Thaler 1980). Kahneman et al. (1990) demonstrate this phenomenon by endowing half of their participants with a coffee mug and then allowing them to transact with the other half who did not receive a mug. Based on the

Coase theorem, approximately fifty percent of the mugs should be sold. However, very few transactions occurred because the amount that owners who were endowed with the mug were willing to accept (WTA) to sell their mugs was more than twice as large as the amount that non-owners were willing to pay (WTP) to buy them. This endowment effect finding that mere ownership of an item causes an individual to value it higher than if it is not owned has been confirmed in many studies (Huck, Kirchsteiger, and Oechssler 2005; Johnson, Häubl, and Keinan 2007; Shu and Peck 2011; Weaver and Frederick 2012; Morewedge and Giblin 2015).

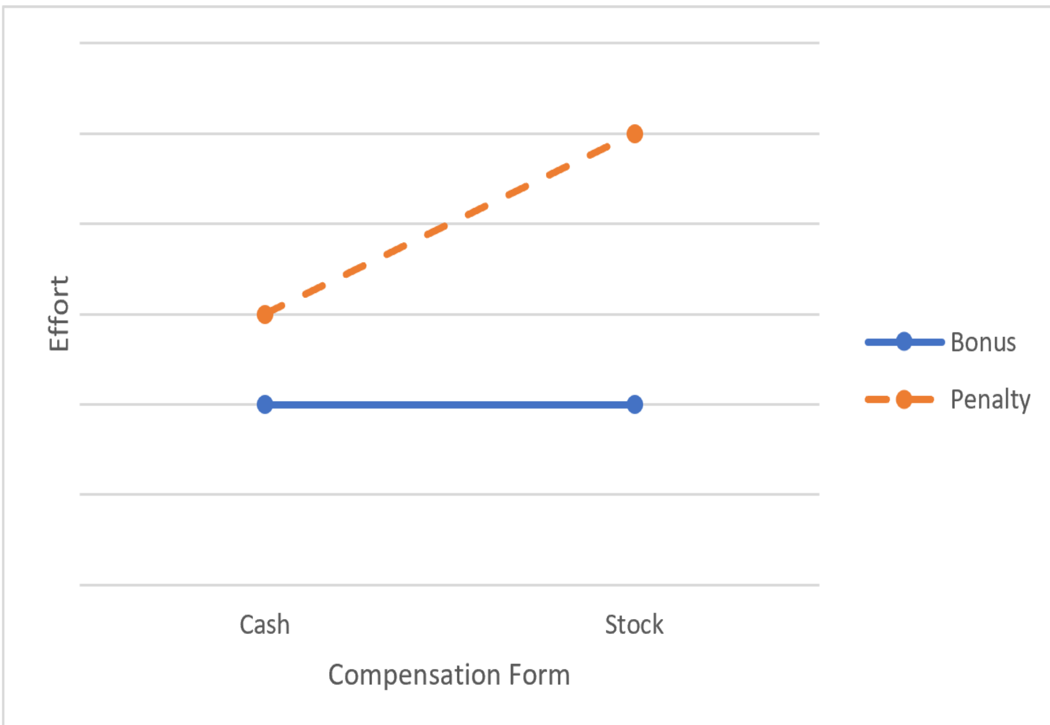
Individuals working under a penalty-framed contract effectively own the penalty portion of their compensation (i.e., it is part of their endowment) and therefore will likely value it more due to the endowment effect. Individuals working under a bonus-framed contract do not own the bonus portion of their compensation, so its value is unaffected by the endowment effect. Thus, I expect individuals will place a higher value on the penalty portion of their compensation under a penalty-framed contract than on the bonus portion of their compensation under a bonus-framed contract. Importantly, Kahneman et al. (1990) identify a critical exception to the endowment effect, in that it does not occur for cash or near-cash items because such items are difficult to mentally re-value. For example, individuals do not re-value $10 of cash to be worth more than $10 whether they own the $10 or not. Kahneman et al. (1990) demonstrate this by endowing half their participants with cash-like induced-value tokens (tokens that have a different assigned value to each participant) rather than mugs. They find that the gap in the amount an owner is willing to accept (WTA) to sell their token and the amount a non-owner is willing to pay (WTP) to buy a token disappears and, as the Coase theorem predicts, fifty percent of all tokens were sold.

Thus, although individuals generally value items they own (as in the case of penalty-framed contracts) more than items they do not own (as in the case of bonus-framed contracts) due to the endowment effect, the valuation of cash is unaffected by ownership. As such, the endowment effect is expected to increase the perceived value of stock in a penalty-framed contract, but not affect the perceived value of cash in a penalty-framed contract. The stock in a penalty-framed contract is also expected to be valued higher than either cash or stock in a bonus-framed contract because the potential bonus amounts in those contracts are not part of an individual's endowment.

In summary, despite economic predictions to the contrary, archival evidence shows that stock is at least as effective an incentive as cash. Two common forms of stock are unrestricted stock and restricted stock with vesting requirements. I predict individuals will perceive unrestricted stock as a bonus and restricted stock with vesting requirements as a penalty. Prior research shows that individuals provide more effort under penalty-framed contracts than bonus-framed contracts. Additionally, due to the endowment effect, individuals are expected to value the penalty portion of their compensation higher, but will not value the economically equivalent bonus portion of their compensation higher. Importantly, this endowment effect applies to stock which can be mentally revalued, but not to cash which is difficult to mentally revalue. Therefore, due to the endowment effect, I expect employees to increase their valuation of stock in a penalty-framed stock contract, but not to increase their valuation of cash in a penalty-framed cash contract. Finally, because individuals work harder for compensation they value more, I predict that effort will be higher under the penalty-framed stock contract, which represents restricted stock arrangements with vesting requirements, than under penalty-framed cash contracts or bonus-framed cash or stock contracts. These predictions are summarized in Figure 1 and formally stated in my hypothesis.

H1: The motivating effect of a penalty frame (as opposed to a bonus frame) on effort will be greater for stock compensation than for cash compensation, such that effort will be higher under a penalty-framed stock contract than under a penalty-framed cash contract or a bonus-framed stock or cash contract.

Figure 1: Hypothesized Patten of Results[a]



[a] See Table 6 for a description of the how I operationalize Cash, Stock, Bonus, and Penalty conditions, and measure Effort.

3.3 Method

3.3.1 Preliminary Survey

Before collecting the main data used to test my hypothesis, I conducted a survey to provide evidence in support of my expectation that employees perceive unrestricted stock as a bonus and restricted stock as a penalty. I provide survey participants with textbook definitions of unrestricted and restricted stock. Then, I ask them to provide their perceptions of two different compensation contracts. In the first contract, participants are told that, in addition to their salary, they will receive shares of unrestricted stock in one year if they perform well and no shares if they do not perform well. In the second contract, participants are told that, in addition to their salary, they will receive shares of restricted stock today that will vest in one year if they perform well, but be forfeited if they do not perform well. After reading the description of each contract, participants responded with their level of agreement on a 7-point scale (strongly disagree [1] to strongly agree [7]) to two statements: 1) The portion of my compensation involving (un)restricted stock feels like a bonus for performing well and 2) The portion of my compensation involving un(restricted) stock feels like a penalty for not performing well.

157 participants from a large university's lab completed the survey and earned course credit for their participation. Participants were an average of 20.0 years old and 53 percent were male. Participants' mean agreement with the aforementioned statements indicate they perceive unrestricted stock to be more like a bonus than restricted stock (5.93 versus 4.87, $t = 6.36$, $p < 0.01$).[19] In addition, they perceive restricted stock to be more like a penalty than unrestricted

[19] All p-values are reported one-tailed due to directional predictions unless noted otherwise.

stock (5.50 versus 4.68, $t = 4.41$, $p < 0.01$). This supports my expectation that employees perceive unrestricted stock contracts as bonus-framed contracts and restricted stock contracts as penalty-framed contracts.[20]

3.3.2 Experimental Procedures

I use an experiment to investigate the effect of vesting requirements on rank-and-file employees' effort. Using an experiment to investigate my research question is advantageous for two important reasons. First, an experiment allows me to isolate the effect that vesting requirements have on employee effort. Prior research has identified other factors that increase the effectiveness of stock compensation in incentivizing effort. For example, Hochberg and Lindsey (2010) find stock-based compensation leads to increased effort due to increased mutual monitoring. In my experiment, I abstract away from factors such as these to isolate the effect of

[20] Although these results provide evidence consistent with employees perceiving unrestricted stock as a bonus and restricted stock as a penalty, participants' agreement with the statement that unrestricted stock (restricted stock) contracts felt like a penalty (bonus) was also above the scale midpoint of 4, albeit to a lesser extent. Therefore, to provide further justification for using bonus-framed stock and penalty-framed stock in my experiment as proxies for unrestricted stock and restricted stock, respectively, I collected additional data that allows me to compare perceptions of these stock contracts to perceptions of the bonus-framed and penalty-framed cash contracts used in the prior bonus versus penalty framing studies. Specifically, I re-ran the survey, but replaced unrestricted stock with a cash incentive payment to be paid in one year for good performance (bonus frame) and replaced restricted stock with a cash incentive payment initially given but to be forfeited without good performance (penalty frame). Sixty-two participants on Amazon's Mechanical Turk completed this second survey. The pattern of results with cash incentives in this second survey is the same as the pattern for stock incentives in the first survey. Participants perceived the bonus-framed cash to be more like a bonus than the penalty-framed cash (5.90 versus 5.27, $t = 2.59$, $p < 0.01$), but still perceived receiving the penalty-framed cash as somewhat of a bonus. Additionally, participants perceived the penalty-framed cash to be more like a penalty than the bonus-framed cash (5.58 versus 4.68, $t = 2.98$, $p < 0.01$), but still perceived not receiving the bonus-framed cash as somewhat of a penalty. Because I find that participants perceive the bonus-framed (penalty-framed) cash contracts used in prior studies the same way they perceived unrestricted stock (restricted stock) contracts, I conclude that it is appropriate to use bonus-framed (penalty-framed) stock contracts as proxies for unrestricted (restricted) stock contracts in my experiment.

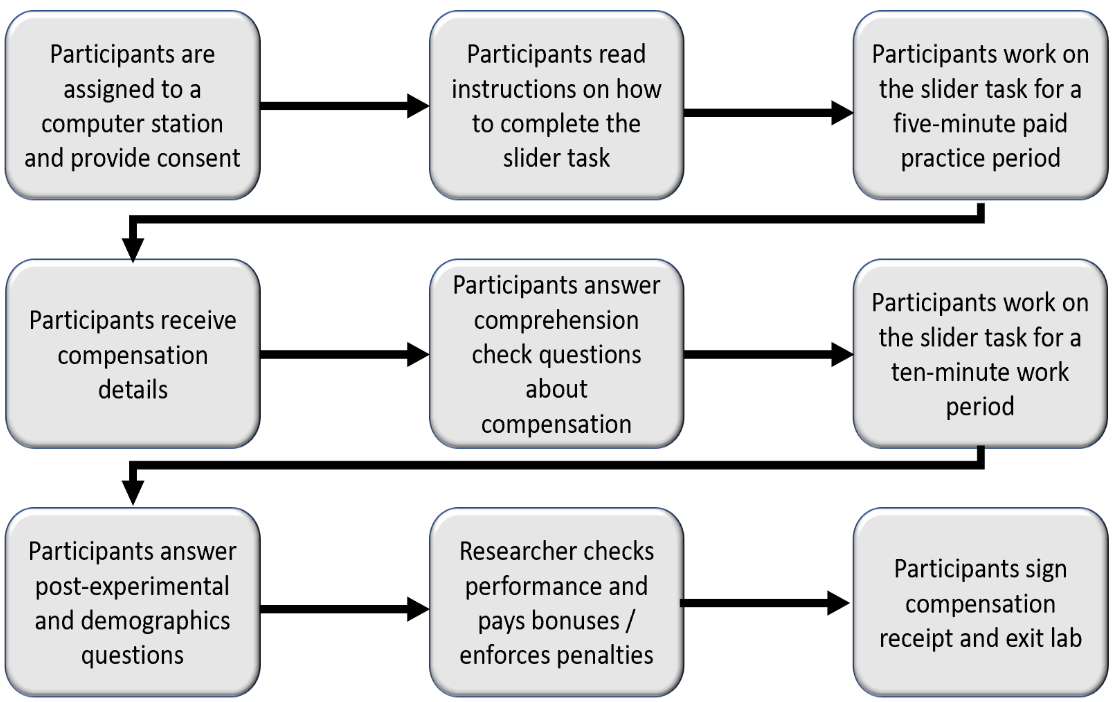
vesting requirements on effort. Second, an experiment allows me to establish a causal link between an employee's compensation and their effort exertion. Core, Guay, and Larcker (2003, p. 34) suggestion caution in interpreting archival studies showing positive performance effects of stock-based compensation, arguing that "a limitation of this research is that the causal direction of the relation between equity incentives and performance is unclear. Rather than higher equity incentives producing better future firm performance, it may be the case that firms expecting better future performance grant more equity." I overcome this limitation by using an experiment.

I use a 2x2 between-participants experimental design in which I manipulate the compensation form (cash versus stock) and the contract frame (bonus versus penalty) between experimental sessions (two sessions per condition for a total of eight sessions). Participants are university students from an experimental economics laboratory participant pool. Students are appropriate to use as participants in my study because they can proxy for the rank-and-file employees that comprise my population of interest (Bonner, Hastie, Sprinkle, and Young 2000; Libby, Bloomfield, and Nelson 2002).

Figure 2 shows an overview of the experimental procedures. After arriving at the lab and providing consent, participants read instructions on how to perform the experimental task, which is an adapted version of Gill and Prowse's (2012) slider task. In this task, numerous sliders are shown on a screen (see Figure 3). Each slider has a horizontal scrollbar with endpoints of 0 and 100. The slider marker is initially positioned at 0, and participants must use their mouse to manually drag the slider marker to the midpoint of the scrollbar (50). Each slider that is completed on the screen counts as one unit of performance. Participants' performance on the slider task is a good proxy for effort because there is a strong link between effort and performance (see the Choi, Clark, and Presslee 2019 paper presented in Chapter 2.0 of this dissertation). Participants worked on the slider task during a five-minute practice period, earning piece-rate pay of $0.02 per slider completed.

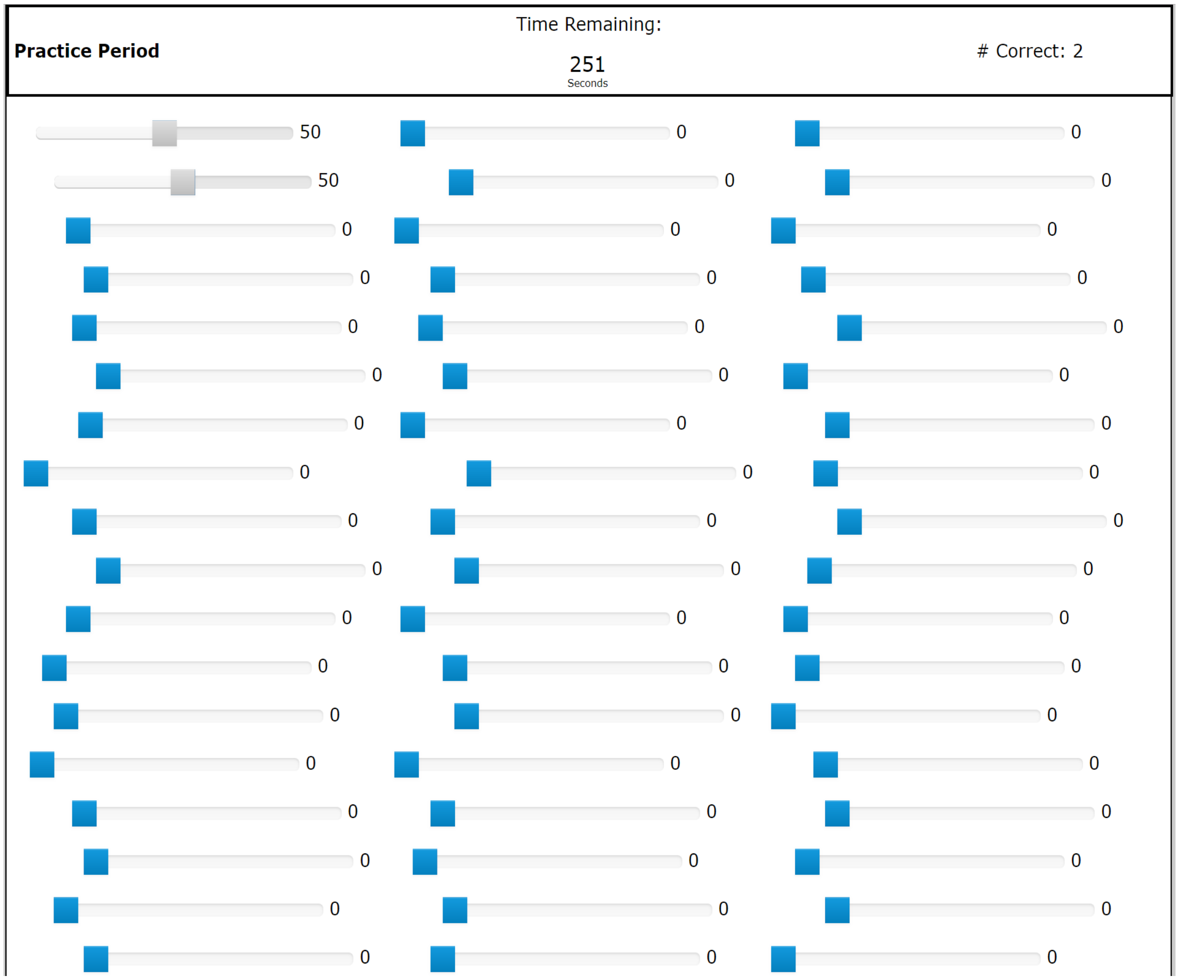
After the practice period, participants assumed the role of an employee in a hypothetical firm working under an assigned compensation contract. Next, participants took a quiz regarding the task and their compensation. Then, they worked on the slider task during a ten-minute work period before responding to the post-experimental questionnaire. Finally, before participants left the lab, bonuses were paid and penalties were collected according to participants' compensation contracts and performance.

Figure 2: Experimental Overview[a]



[a] Figure 2 presents an overview of the experimental procedures followed in the lab.

Figure 3: Slider Task[a]



[a] In the slider task, participants must drag the slider markers, initially positioned at the "0" position, to the "50" position. Each correctly positioned slider marker counts as one unit of performance. In the example shown here, the first two sliders in the first column have been completed.

3.3.3 Compensation Form

All participants received \$7, and I manipulate whether they earn an additional \$5 or a share of stock with a current value of \$5 by achieving a performance target of 180 sliders.[21] The value of the share of stock was directly tied to the stock price of a real company and was worth \$5 on the date of the experiment. I informed participants that I would mail them the monetary value of their potential share of stock four weeks later, and that the value on that date would depend on the stock price of the real company (e.g., if the actual stock price changed by 5 percent, the value of participants' stock also changed by 5 percent in the same direction). I also told participants I would email them the name of the company at the conclusion of all of the experimental sessions so they could track the value of their stock. But, at the time of the experiment, the only information they had regarding the company was that it was a large employer of recent graduates from their university.

Despite the differences between the cash and stock contracts, finance theory indicates that they are economically equivalent. On the date of the experiment, when participants decide how hard to work, both are worth \$5. The two differences between the two compensation forms are 1) the cash is paid to participants on the day of the experiment whereas the stock payment comes four weeks later, and 2) the value of the stock is variable, causing it to be inherently riskier than cash. In finance theory, the capital asset pricing model, or CAPM, is commonly used

---

[21] The performance target is set to the median performance of a separate group of students who completed the task previously. Thus, approximately fifty percent of participants will reach the target. I chose this moderately difficult goal to encourage all participants to work towards meeting the target, thereby earning the bonus or avoiding the penalty in their compensation contract. If the performance target was too difficult (e.g., only 20 percent chance of attainment), I would likely observe quitting behavior whereby participants do not exert effort due to a low perceived likelihood of achieving the performance target, even if participants differentially valued their compensation (Locke and Latham 1990). However, if the performance target was too easy (e.g., 80 percent chance of attainment) participants may exert minimal effort because they feel assured of achieving the performance target.

to calculate the theoretical rate of return for an asset (Sharpe 1964). According to this model, investors need to be compensated for both the time value of money and for bearing risk. So, although participants in stock conditions receive payment later and bear more risk than participants in the cash conditions, they are compensated for this by the expected increase in the value of their real company's stock.

3.3.4 Contract Frame

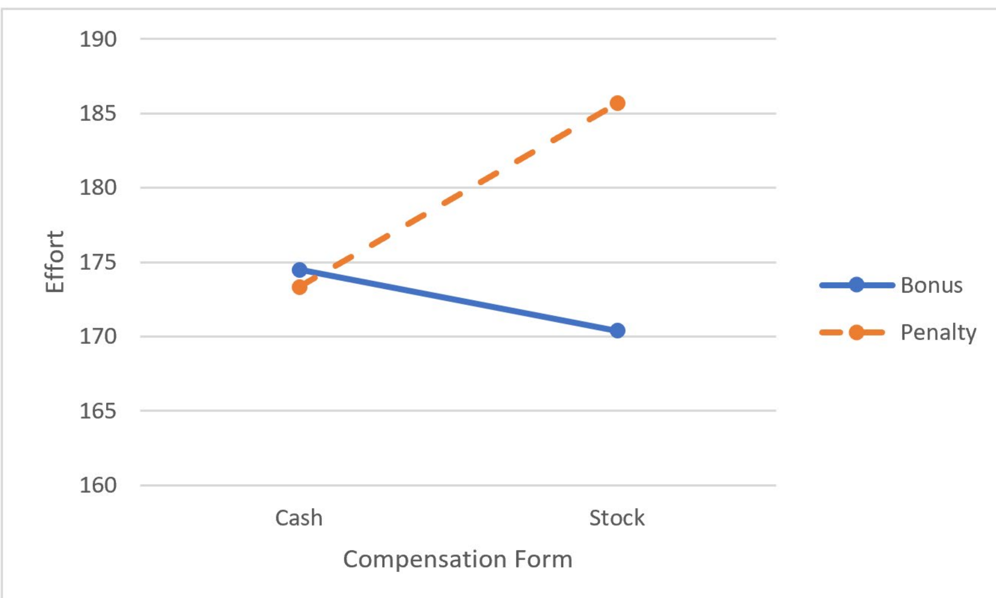
I manipulate the frame of the compensation contract to be either a bonus or a penalty. Participants in the bonus-framed conditions learn their compensation is $7 and that if they complete 180 sliders, they will receive a bonus of $5 or a share of stock with a value of $5 on that date. Participants in the penalty-framed conditions learn their compensation is $12 ($7 and a $5 share of stock) but they will forfeit $5 (the $5 share of stock) from their compensation if they do not complete 180 sliders. In addition, after learning about their compensation contract, I asked participants to open a sealed envelope at their work station. For participants in the bonus-framed conditions, the envelope contained $7. For participants in the penalty-framed conditions, the envelope contained either $12 or $7 and a stock certificate with a value of $5.

3.4 Results

156 students participated in my experiment. 92 percent of participants correctly answered comprehension check questions regarding their compensation contract. Of the 12 participants who answered incorrectly, 9 further demonstrated their lack of understanding with irrational

responses in the post-experimental questionnaire, so I exclude these latter 9 participants from the analyses.[22] I use the remaining 147 participants for my analyses throughout the paper. Participants were an average of 20.1 years old, and 64 percent were female. I measure my main dependent variable, effort, as the number of sliders completed during the ten-minute work period. Effort is uncorrelated with age (two-tailed $p = 0.43$), but is correlated with gender ($r = 0.29$, two-tailed $p < 0.01$) such that male participants performed better than female participants. However, I do not control for gender because doing so does not affect any of my statistical inferences. I report descriptive statistics related to my hypothesis in Table 6, Panel A. I present a graphical depiction of the results for effort by condition in Figure 4.

Figure 4: Mean Effort by Condition[a]



[a] See Table 6 for a description of the how I operationalize Cash, Stock, Bonus, and Penalty conditions, and measure Effort. The means presented in the graph are covariate-adjusted for ability.

[22] Excluding these participants increases the internal validity of my study by ensuring I only use data from participants who correctly understood their compensation contract. My statistical inferences are unchanged by the decision to exclude these 9 participants, with one exception, which is noted. If the other 3 participants who incorrectly answered comprehension check questions (but responded rationally to the post-experimental questionnaire) were excluded, statistical inferences are unchanged.

3.4.1 Test of H1

As seen in Figure 1, H1 predicts an interaction such that the motivating effect of a penalty frame on effort will be greater for stock compensation than cash compensation, resulting in higher effort under a penalty-framed stock contract than any of the other contracts. To test this prediction, I use an analysis of covariance (ANCOVA) with effort, measured as the number of sliders completed during the ten-minute work period, as my dependent variable, and compensation form (cash or stock), contract frame (bonus or penalty), and their interaction as independent variables. I also include ability, measured as the number of sliders completed during the five-minute practice period, as a covariate. By controlling for ability, I help ensure that performance on the slider task is a good proxy for effort (i.e., differences in the number of completed sliders are driven by differences in effort exertion and not differences in ability). Indeed, I find my measure of ability is significantly correlated with the number of completed sliders during the ten-minute work period ($r = 0.54$, two-tailed $p < 0.01$). Therefore, controlling for ability removes noise from the data, allowing me to isolate the effect of my independent variables on effort.[23]

I report the results of the ANCOVA in Table 6, Panel B. Consistent with H1, I find a statistically significant interaction of contract frame and compensation form, such that the effect of contract frame on effort differs for cash and stock compensation ($F = 5.89$, $p = 0.01$). As a result, I find stock compensation leads to greater effort than cash compensation in penalty-

[23] Controlling for ability is particularly important in this study because randomization failed to equally distribute ability across all conditions. Mean ability in the bonus-framed stock condition (84.69) is significantly greater than the mean ability in the other three conditions combined (77.02; $t = 2.09$, two-tailed $p = 0.04$).

framed contracts ($t = 2.63$, $p < 0.01$), but not in bonus-framed contracts ($t = -0.84$, two-tailed $p = 0.40$; See Table 6, Panel C).

I also perform two additional tests to examine the specific pattern I predict in H1 (Figure 1). First, I investigate whether effort is highest under the penalty-framed stock contract due to the combined effects of loss aversion and the endowment effect. I find support for this prediction, with effort under the penalty-framed stock contract (185.68) being higher than effort under the penalty-framed cash contract (173.33, $t = 2.63$, $p < 0.01$), the bonus-framed cash contract (174.50, $t = 2.30$, $p = 0.01$), and the bonus-framed stock contract (170.39, $t = 2.98$, $p < 0.01$).

Second, recall that I predict penalty-framed cash contracts lead to more effort than bonus-framed cash contracts due to loss aversion. Unexpectedly, as reported in Table 6, Panel C, I find a penalty frame does not lead to greater effort than a bonus frame for cash compensation ($t = -0.26$, two-tailed $p = 0.80$).[24] Thus, at first glance, my results seem inconsistent with Hannan et al.'s (2005) results.

Hannan et al. (2005) find penalty-framed contracts lead to reduced feelings of fairness relative to bonus-framed contracts (leading to reduced effort), and penalty-framed contracts lead to increased feelings of disappointment (loss aversion) relative to bonus-framed contracts (leading to increased effort). In a path model, Hannan et al. (2005) show both fairness and disappointment mediate the effect of contract frame on effort, but a penalty-framed contract leads to greater effort because the effect of disappointment dominates the effect of fairness. As reported in Figure 5, I replicate the majority of their path model, with the exception that the path

[24] I report the results of this test using a two-tailed p-value, despite my directional prediction, because the result is in the opposite direction of my prediction.

between disappointment and effort only approaches marginal significance (p = 0.13).[25] However, unlike in Hannan et al.'s (2005) study, the relative effects of fairness and disappointment on effort appear to be more balanced in my study, causing the two effects to fully offset each other.

Table 6: Descriptive Statistics and Results

Panel A: Mean (Standard Deviation) by Condition[a]

	Cash		Stock	
	Bonus	**Penalty**	**Bonus**	**Penalty**
n	38	45	32	32
Ability	78.66 (19.74)	75.78 (22.16)	84.69 (15.44)	76.81 (13.18)
Effort	174.47 (25.64)	171.16 (22.52)	174.88 (27.51)	184.28 (22.45)
Effort (Covariate-Adjusted for Ability)	174.50 (20.08)	173.33 (16.69)	170.39 (24.07)	185.68 (20.77)

[25] Hannan et al. (2005) use a probabilistic task, such that when participants responded to a post-experimental question about their expected disappointment, they were still unaware of whether they had earned the bonus or avoided the penalty. In my study, however, participants received real-time feedback about their performance, and thus knew whether they had reached the performance target when responding to the post-experimental questionnaire. As such, I asked participants who reached (didn't reach) the performance target "How disappointed would you have been if you had not achieved your performance target and therefore had not received $5 as a bonus?" (How disappointed are you that you did not achieve your performance target and therefore did not receive $5 as a bonus?)" It is likely that Hannan et al.'s (2005) measure of participants' expected disappointment is more predictive of behavior than my outcome-based measure. For example, a participant may not believe that he will experience disappointment if he misses out on a bonus and therefore, does not work hard. However, contrary to his expectation, he feels very disappointed when failing to earn the bonus, but at this point there is no opportunity to exert more effort. In situations such as these, my measure of disappointment will not be as predictive of behavior as the one in Hannan et al.'s (2005) study, which may explain why I did not perfectly replicate this path.
.

Panel B: ANCOVA (Dependent Variable: Effort)

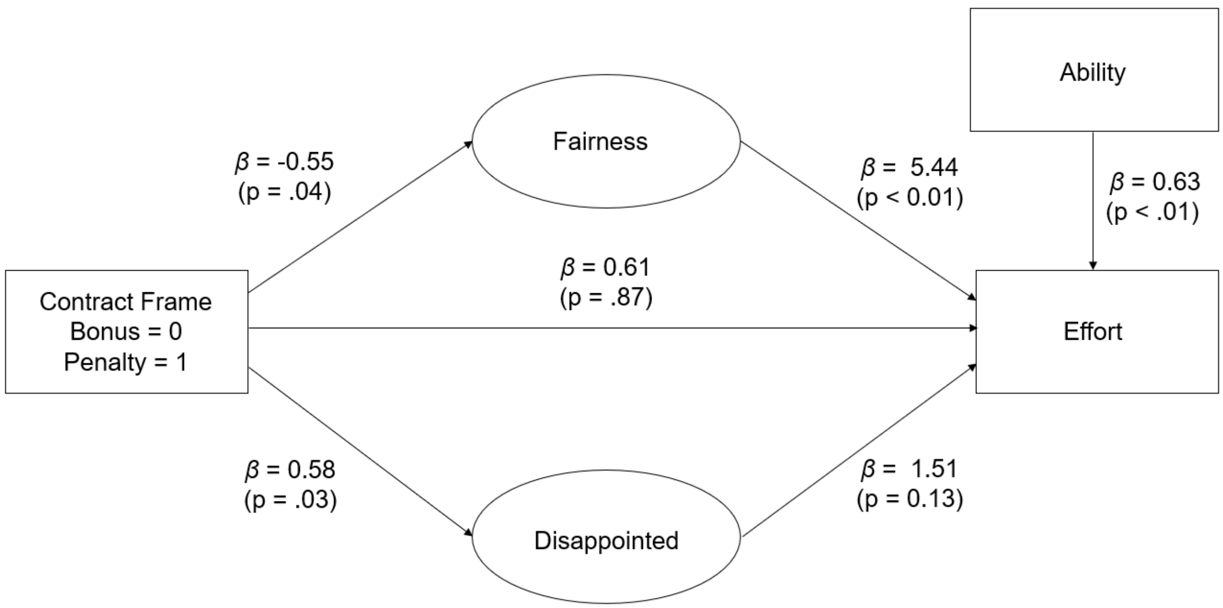
Source	df	MS	F-stat	p-value[b]
Compensation Form	1	607.60	1.47	0.12
Contract Frame	1	1,757.97	4.27	0.02
Compensation Form x Contract Frame	1	2,427.17	5.89	0.01
Ability	1	27,220.06	66.07	< 0.01
Error	142	412.0		
Adjusted R^2 = 32.48%				

Panel C: Simple Effects Tests

Source	df	t-stat	p-value[b]
Bonus versus Penalty in Cash	1	-0.26	0.80, two-tailed
Bonus versus Penalty in Stock	1	2.98	< 0.01
Cash versus Stock in Bonus	1	-0.84	0.40, two-tailed
Cash versus Stock in Penalty	1	2.63	< 0.01

[a] 147 participants were randomized into one of four conditions. *Compensation Form* (Cash = 0 vs. Stock = 1) and *Contract Frame* (Bonus = 0 vs. Penalty = 1) are manipulated between-sessions. In *Cash* conditions, the bonus or penalty was $5. In *Stock* conditions, the bonus of penalty was a share of stock with a current value of $5. The value was tied to the stock price of an actual company and participants would receive its value four weeks later. In *Bonus* conditions, participants were told that if they reach their performance target, they would receive a bonus. In *Penalty* conditions, participants were initially given their form of compensation and they would lose it as a penalty if they did not reach their performance target. Ability (Effort) is measured by the number of sliders completed during the five-minute practice period (ten-minute work period).
[b] All p-values are reported one-tailed due to directional predictions unless noted otherwise.

Figure 5: Replication of Hannan et al. (2005)[a]



[a] This analysis only looks at participants in the cash compensation conditions. 83 participants were randomized to either the bonus-framed cash condition (38) or the penalty-framed cash condition (45). See Table 6 for a description of *Contract Frame, Effort*, and *Ability*. *Fairness* is measured as participants' response on a 7-point Likert scale to the post-experimental question: "How fair did you consider the compensation contract used in this study?" *Disappointment* is measured as participants' response on a 7-point Likert scale to the post-experimental questions "How disappointed would you have been if you had not achieved your performance target and therefore had not received $5 as a bonus" or "How disappointed are you that you did not achieve your performance target and therefore did not receive $5 as a bonus," depending on whether or not the participants reached their performance target.

The theory underlying H1 asserts participants working under a penalty-framed contract will feel as if they are endowed with their compensation and subsequently assign a greater value to it. However, this will only occur with stock compensation because the endowment effect does not occur for cash. In the stock conditions, I measure participants' valuations of the stock by asking them to "please give your best estimate of the value of the share of stock in four weeks." Despite having identical information regarding the stock, participants in the penalty-framed stock contract condition (who were endowed with the stock) estimate its value to be greater than participants in the bonus-framed stock contract condition (who were not endowed with the stock) (\$8.13 versus \$6.05, $t = 2.00$, $p = 0.03$). That is, the penalty-framed stock contract, which represents restricted stock with vesting requirements, caused participants to assign an approximately one-third greater value to their stock compensation than the bonus-framed stock contract, which represents unrestricted stock.

Because participants know whether or not they reached their performance target before completing the post-experimental questionnaire, an alternative explanation for this result is that participants who reach their performance target evaluate the stock in a biased manner and are more likely to estimate a higher value. This may occur because loss aversion predicts participants are more likely to reach their performance targets in penalty-framed stock contracts than bonus-framed stock contracts. To rule out this alternative explanation, I conduct an OLS regression with the estimated stock price as the dependent variable, contract frame (1=penalty, 0=bonus) as the independent variable, and whether a participant reached the performance target as a control variable (1 = reached target, 0 = missed target). Untabulated results indicate the effect of contract frame on estimated stock price remains significant ($\beta = 1.54$, $t = 1.71$, $p = 0.05$) when controlling for whether or not participants reached their target.[26] This is consistent with my theory that the higher valuations of stock under penalty-framed contracts than under bonus-framed contracts were caused by the endowment effect.

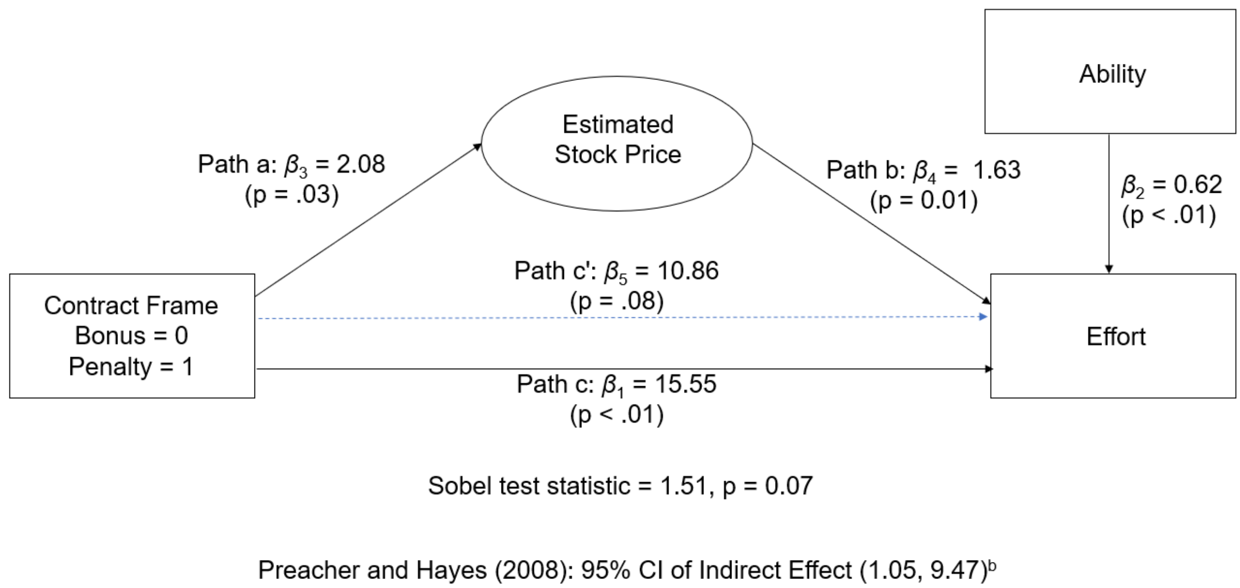
Following Baron and Kenny (1986), I test whether the estimated stock value mediates the effect of contract frame on effort. I do not ask participants how much they value the \$5 payment

---

[26] In addition, the size of the effect of contract frame on estimated stock price is not statistically different after controlling for incentive attainment. That is, in an OLS regression of estimated stock price on contract frame, the coefficient of contract frame ($\beta = 2.08$) is not significantly different than the coefficient of contract frame ($\beta = 1.54$) when performance target attainment is included in the model ($\chi^2 = 0.98$, $p = 0.32$).

in the cash conditions, as I expect they would all respond with $5, showing no difference in endowment (Kahneman et al. 1990). Thus, I test for mediation only within the stock conditions. I present a graphical representation of the mediation analysis in Figure 6.

Figure 6: Tests of Mediation[a]



[a] See Table 6 for a description of *Contract Frame, Effort,* and *Ability. Estimated Stock Price* is measured as participants' response to the post-experimental item: "Please give your best estimate of the value of the share of stock in four weeks." The coefficients and p-values shown are from the following regressions:
$Effort = \alpha + \beta_1 ContractFrame + \beta_2 Ability + \varepsilon$
$EstimatedStockPrice = \alpha + \beta_3 ContractFrame + \varepsilon Effort = \alpha + \beta_4 CEstimatedStockPrice + \beta_5 ContractFrame + \beta_2 Ability + \varepsilon$
[b] I estimate the 95 percent confidence interval of the indirect effect of *EstimatedStockPrice* on *Effort* using 5,000 bootstrapped samples (Preacher and Hayes 2008). The indirect effect is significant if the confidence interval does not include zero, which suggests that mediation occurred.

Mediation occurs if the following four conditions are satisfied: (1) a significant relationship exists between the independent variable, contract frame, and the dependent variable, effort (path c), (2) a significant relationship exists between the independent variable and the mediator, estimated stock price (path a), (3) a significant relationship exists between the mediator and the dependent variable (path b), and (4) a decrease in the significance of the relationship between the independent and dependent variables after controlling for the mediator (path c′). I continue to control for the effect of ability on effort during my tests of mediation.

An indicator variable for contract frame (bonus = 0, penalty = 1) has a positive effect on effort ($t = 2.64$, $p = 0.01$) so condition (1) is satisfied (path c). Contract frame has a positive effect on the estimated stock price ($t = 2.82$, $p = 0.01$) so condition (2) is satisfied (path a). Estimated stock price has a positive effect on effort ($t = 2.29$, $p = 0.03$) so condition (3) is satisfied (path b). Finally, the effect of contract frame on effort decreases in statistical significance when controlling for the estimated stock price ($t = 1.80$, $p = 0.08$) but the effect remains marginally statistically significant (path c′). A Sobel test (Sobel 1982) indicates the decrease in explanatory power of contract frame on effort is marginally statistically significant (Sobel test statistic = 1.78, $p = 0.07$). These results suggest the effect of contract frame on effort is partially mediated by the estimated stock price, which is higher for penalty contracts due to the endowment effect.[27, 28]

3.4.2 Supplemental Analyses

As discussed previously, economic theory predicts stock compensation is a less effective incentive than cash compensation. However, the available archival evidence finds stock is at

---

[27] I also test whether estimated stock price mediates the relationship between contract frame and effort using Preacher and Hayes' (2008) method. I estimate the 95 percent confidence interval of the indirect effect of the estimated stock price on effort to be (1.05, 9.47) using 5,000 bootstrapped samples. Since the confidence interval does not contain zero, I conclude estimated stock price mediates the relationship between contract frame and effort.

[28] As mentioned previously, I excluded the data from nine participants who failed manipulation checks. These participants provided extremely high and unreasonable estimates of the future stock price, ranging from $50 to $1,000. If these individuals are included, the mean (standard deviation) estimated stock price is $22.63 (75.31) in the bonus-framed stock condition and $45.68 (165.37) in the penalty-framed stock condition. Additionally, estimated stock price no longer mediates the relationship between contract frame and effort.

least as effective an incentive as cash. Researchers have theorized that the surprising effectiveness of stock in motivating effort can be attributed to various factors associated with stock compensation, such as improved employee selection and retention and an increase in mutual monitoring (Ittner et al. 2003; Hochberg and Lindsey 2010; Hales et al. 2015). In my experiment, I abstract away from factors such as these that could boost the incentive effects of stock. Despite this, I predict and find that stock with vesting requirements is a particularly effective incentive. However, also find no difference in the effectiveness of bonus-framed cash and stock without vesting requirements in motivating effort (t = -0.84, two-tailed p = 0.40). This is mildly surprising since I have abstracted away from the previously-identified mechanisms that explain why stock is at least as motivating as cash. So, in this supplemental analysis, I investigate if optimism bias can explain why bonus-framed stock and cash contracts led to similar levels of effort in my study.

Optimism bias reflects the tendency for individuals to overestimate the likelihood of positive events and to underestimate the likelihood of negative events (Weinstein 1980). Researchers have refined our understanding of optimism bias by identifying sub-categories of optimism. Specifically, *dispositional* optimism refers to a general belief that good outcomes will occur (Scheier and Carver 1985, 1994). In contrast, *situational* optimism (Armor and Taylor 1998; Kluemper, Little, and DeGroot 2009) is how optimistic an individual feels in a specific situation. Unlike *dispositional* optimism which is a stable trait, *situational* optimism varies by setting. Thus, even individuals with low *dispositional* optimism (i.e., *dispositional* pessimists) can be *situationally* optimistic in certain settings. In settings where *dispositional* and *situational* optimism levels differ, research suggests *situational* optimism is a better predictor of psychological and biological responses (Taylor et al. 1992; Nonis and Wright 2003). Thus, *dispositional* pessimists who are *situationally* optimistic about the stock market would respond in an optimistic way to stock compensation.

Anecdotal evidence suggests individuals are, on average, *situationally* optimistic about the stock market, even when the underlying fundamentals suggest such optimism is unwarranted. Several prominent economists have remarked that investors are overly optimistic regarding the stock market. For example, Nobel laureate Robert Shiller has repeatedly claimed investors are irrationally exuberant (Shiller 2001, 2016.) This claim is supported by a measure of market optimism he created, the cyclically adjusted price-earnings ratio, which suggests that the market has experienced elevated optimism for the last 25 years (Shiller 2018).

There are various potential explanations for <u>why</u> investors are situationally optimistic regarding the stock market, making it difficult to identify the exact reason for the optimism. One possible reason for high levels of *situational* optimism about the stock market could be the use of counterfactual reasoning, which is a mental representation of an alternative outcome to the outcome experienced. Individuals commonly create upward counterfactuals of better possible outcomes after experiencing a negative event, but do not think of downward counterfactuals after a positive event (Roese 1997). This way of thinking could cause individuals who experience bad outcomes in the stock market to create upward counterfactuals such as “if only I had made a different investment, I would have earned a great return.” However, individuals who experienced a good outcome are less likely to create downward counterfactuals such as “if I had not made that investment, I would have missed out on a great return.” This tendency would cause individuals to discount the likelihood of having a bad outcome in the stock market, which could contribute to the elevated level of *situational* optimism in this setting. Given this background on optimism bias, I next investigate whether participants were situationally optimistic with regard to their stock's value in my study, enhancing its motivational effect and causing bonus-framed stock and cash compensation to lead to similar levels of effort despite economic theories predicting otherwise.

Because situational optimism involves expectations about outcomes in specific settings, measures of situational optimism must be different for different settings. To my knowledge, no measure exists on situational optimism related to the stock market, so I needed to create one. To do so, I generated a list of items by considering stock-related outcomes, as well as factors that might contribute to stock evaluations (e.g., creating upward, but not downward counterfactuals and the availability of positive versus negative stock-related experiences). Higher responses for the items in my list indicate a greater degree of situational optimism relating to the stock market. I present the list of the six items in Table 7, Panel A. To test whether the items measure a single construct, I perform exploratory factor analysis (EFA). This analysis produces one factor with an eigenvalue of 1.23, which explains 58.76 percent of the variance. The rotated factor loadings, as reported in Table 7, Panel B, indicate this factor is composed of two items that capture optimism about the stock market as a whole (*OverallStock* and *StockOptimism*). The rotated factor loadings for the two items are 0.76 and 0.77, respectively, and Cronbach's alpha is 0.76. Thus, I average the responses for these two items to create a single factor, which I label *Situational Optimism*.

To help validate this factor as separate from dispositional optimism, I also measure *Dispositional Optimism* using the Life Orientation Test-Revised (LOT-R), an established scale from psychology (Scheier, Carver, and Bridges 1994). I find *Situational Optimism* and *Dispositional Optimism* are not correlated ($r = 0.03$, two-tailed $p = 0.82$), consistent with prior research in the area.[29] Prior research also finds situational optimism is a better predictor of behavior than dispositional optimism. I test whether this is true in my setting.

If participants are optimistic with regard to the stock market, I expect they will overvalue their stock compensation relative to an economic benchmark. To test this, I return to the previously referenced post-experimental question that asks participants to estimate the stock price in four weeks. Note that the endowment effect does not affect the stock price estimates in the bonus-framed stock condition I am analyzing because participants working under that contract did not own their stock. The value of the share of stock on the date of the experiment was \$5. The mean (standard deviation) of the estimated value of the bonus-framed stock was \$6.05 (3.36). To be able to compare the estimated stock value of \$6.05 in four weeks to the \$5 that participants in the cash conditions could earn immediately, I calculate the net present value of the estimate assuming a 10 percent annual interest rate. The result, \$6.00, is marginally greater than the \$5 participants in the cash compensation conditions could earn ($t = 1.70$, two-tailed $p = 0.10$). The high return implied by participants' estimates suggests participants have high levels of optimism with regard to the stock market. This is consistent with Hodge, Rajgopal, and Shevlin (2009), who also find individuals overvalue stock-based compensation relative to an economic benchmark.

[29] Although research on situational optimism has typically found it is not correlated with dispositional optimism, this is likely an artifact of which settings have been studied. That is, it is likely that dispositional and situational optimism are correlated in most settings, but situational optimism has been studied in settings where researchers believe it will differ from an individual's dispositional optimism.

Table 7: Situational Optimism[a]

Panel A: Measures

1) How likely do you think it is that the price of your stock will increase between now and when the stock payout happens in four weeks? [*SpecificStock*]

   | --------------------------------| ___%
   0% 100%

2) I believe that the value of the stock market, as a whole, will increase over time. [*OverallStock*]

   1 2 3 4 5 6 7
   |--------------|-------------|---------------|--------------|-------------|-------------|
   Strongly disagree — Neither agree nor disagree — Strongly agree

3) Imagine you bought stock in a company and the value decreased over time. It is likely that if you had bought stock in a different company, the value would have increased over time. [*PositiveCounterfactual*]

   1 2 3 4 5 6 7
   |--------------|-------------|---------------|--------------|-------------|-------------|
   Strongly disagree — Neither agree nor disagree — Strongly agree

4) Imagine you bought stock in a company and the value increased over time. It is likely that if you had bought stock in a different company, the value would have decreased over time. [*NegativeCounterfactual*]

   7 6 5 4 3 2 1
   |--------------|-------------|---------------|--------------|-------------|-------------|
   Strongly disagree — Neither agree nor disagree — Strongly agree

5) When you hear other people talk about their investments in the stock market, do they generally talk more about positive experiences or more about negative experiences? [*StockTalk*]

   1 2 3 4 5 6 7
   |--------------|-------------|---------------|--------------|-------------|-------------|
   Negative experiences — Equal positive and negative — Positive experiences

6) I am optimistic about the stock market. [*StockOptimism*]

   1 2 3 4 5 6 7
   |--------------|-------------|---------------|--------------|-------------|-------------|
   Strongly disagree — Neither agree nor disagree — Strongly agree

Panel B: Factor Analysis[b]

Measure	Rotated Factor Loading
SpecificStock	-0.01
OverallStock	0.76
PositiveCounterfactual	0.00
NegativeCounterfactual	0.05
StockTalk	-0.04
StockOptimism	0.77

[a] The 64 participants in *Stock* conditions completed an assessment of Situational Optimism in the post-experimental questionnaire.
[b] Exploratory factor analysis (EFA) produces one factor with an eigenvalue of 1.23, which explains 58.76 percent of the variance. The factor is comprised of *OverallStock* and *StockOptimism*, which have a Cronbach's alpha of 0.76.

Further, I demonstrate that this overvaluation is driven by situational optimism and not dispositional optimism. Specifically, I conduct an OLS regression with the estimated stock price as the dependent variable, and *Situational Optimism* and *Dispositional Optimism* as independent variables. Untabulated results indicate *Situational Optimism* is positively associated with the estimated stock price ($\beta = 1.19$, $t = 2.44$, $p = 0.01$), but *Dispositional Optimism* is not ($\beta = 0.34$, $t = 0.49$, $p = 0.63$). Additionally, since participants who value their stock more highly should work harder to earn it, I also expect *Situational Optimism* to be associated with participants' effort. I conduct an OLS regression with effort as the dependent variable, and *Situational Optimism* and *Dispositional Optimism* as independent variables. Untabulated results indicate *Situational Optimism* is positively associated with effort ($\beta = 11.68$, $t = 3.10$, $p < 0.01$), but

Dispositional Optimism is not (β = 6.79, t = 1.27, two-tailed p = 0.21).[30] This set of results suggests participants felt situationally optimistic towards the stock market, causing them to assign more value to their stock and subsequently work harder to earn it. This helps explain why I find bonus-framed stock and cash compensation lead to similar levels of effort despite 1) economic predictions that stock is an ineffective incentive and 2) abstracting away from previously-identified mechanics that increase the incentive effects of stock.

Finally, I explore the effects of risk aversion in my setting. Research shows individuals are risk-averse, on average, which makes stock compensation less attractive. I measure risk aversion using a task adapted from Holt and Laury (2002; see Appendix D). Inconsistent with economic theory, but consistent with Hodge et al. (2009), I find risk aversion is uncorrelated with estimates of stock value (r = -0.13, two-tailed p = 0.29) and effort (r = -0.05, two-tailed p = 0.56), suggesting risk aversion did not play a significant role in my setting.

3.5 Conclusion

This study investigates whether restricted stock with a vesting requirement leads to more effort than other economically equivalent cash and stock compensation arrangements. In a survey, I find individuals perceive restricted stock with a vesting requirement as a penalty and

[30] This is consistent with evidence from Hales et al. (2015). To test their theory, they intentionally use a task with a weak link between effort and performance and find individuals who have higher levels of dispositional optimism work harder because they are more optimistic that their effort exertion will be successful in improving their performance. However, when they analyze the subsample of their participants who are experienced with the task (i.e., there is less uncertainty between effort and performance), they find dispositional optimism is no longer associated with higher performance. This latter result is relevant to my study because there is a strong link between effort and performance in my experimental task (Bonner and Sprinkle 2002).

unrestricted stock without a vesting requirement as a bonus. Then in my experiment, I find effort is higher under a penalty-framed stock contract than under a penalty-framed cash contract or a bonus-framed cash or stock contract. This result occurs because of the endowment effect, under which the motivating effect of a penalty frame on effort is stronger for stock than for cash.

My study contributes to the literature on incentives by demonstrating that stock contracts with features comparable to vesting requirements lead to more effort than other types of economically equivalent contracts. This helps explain why restricted stock is so prevalent in rank-and-file compensation contracts. My study also contributes to the literature on contract frames by showing penalty-framed contracts effectively endow employees with the penalty portion of a contract's compensation. When the compensation form is not cash, I find participants place greater subjective value on this portion of the compensation due to the endowment effect. This insight was not previously discovered because studies on bonus- and penalty-framed contracts used cash as the compensation form. Although I test this theory only with stock compensation, this insight potentially extends to other forms of non-cash compensation. For example, providing benefits such as a company cell phone or a designated parking spot in a penalty-framed contract may lead to more effort than providing an equivalent amount of cash in a penalty-framed contract. Finally, my study contributes to the literature on stock compensation by investigating how individuals perceive stock compensation without vesting requirements. I find individuals are situationally optimistic about the stock market, which causes them to overvalue stock compensation relative to an economic benchmark and subsequently work harder to earn that stock.

The limitations of my study provide opportunities for future research. First, individuals' reactions to stock compensation may depend on prevailing market conditions. It is possible that stock valuations would be different in bull versus bear markets, causing differences in how individuals respond to stock compensation.[31, 32] Relatedly, investors may react differently to long- and short-term investments. In my experiment, participants were "invested" in the market for only four weeks, which is for a shorter time duration than for most investments. Also, although I find penalty-framed stock contracts lead to the most effort, firms are limited in the proportion of employee compensation that can have restrictions on its use due to liquidity concerns. The proportion of restricted compensation to overall compensation was relatively high in my setting, and this elevated level may exceed practical levels for rank-and-file employees. Reducing the proportion of the bonus/penalty compensation may reduce the magnitude of differences in effort between the contract types. Finally, List (2003) finds significant experience eliminates the consequences of the endowment effect. If individuals continue to be compensated with stock, their additional experience may limit their increases in valuation due to the endowment effect, which would reduce the increased effort found with penalty-framed stock contracts relative to other contracts.

[31] However, I see no theoretical reason why this and related factors would interact with contract frame. So, the impact of these limitations is likely limited to the form of the interaction.
[32] My experiment took place in April 2018, a month in which stock markets experienced moderate volatility and small positive returns (Financial Engines 2018). No major market fluctuations occurred immediately prior to my experiment.

Appendix A - Relative Merits of Using Real-Effort Tasks

In this appendix, we discuss the relative merits of using real-effort tasks in laboratory experiments. A large body of accounting research uses laboratory experiments to investigate various compensation policy features and their effect on individual effort (Bonner et al. 2000; Bonner and Sprinkle 2002). Of particular interest is the effect of incentive pay on effort, relative to the effects of fixed pay on effort. Many of these studies use a real-effort task whereby participants perform by exerting effort in a specific direction for a set amount of time, and participants' task performance serves as a proxy for their effort.

However, using a real-effort task is not the only method for studying effort. Other laboratory studies use a choice-effort task whereby participants select an effort level using a scale provided by the researcher (e.g., Kuang and Moser 2009). To capture the personal cost of exerting effort, studies using choice-effort tasks specify a function that maps each effort level to a corresponding personal cost incurred by the participant. Participants are aware of the effort cost function before selecting an effort level.

Only a few studies compare inferences using real-effort versus choice-effort tasks, and those that do report evidence suggesting the two types of tasks lead to similar results. For example, Dutcher, Salmon, and Saral (2015) find both types of tasks lead to similar inferences regarding effort in a public goods setting, and Brüggen and Strobel (2007) report similar findings between the two tasks within a gift-exchange setting.

Despite these findings, using a choice-effort task can be advantageous. First, a choice-effort task allows researchers to maintain greater control over experimental parameters. Specifically, a choice-effort task ensures every participant faces an identical effort cost function, while the effort cost function may vary endogenously across participants when a real-effort task is used. Moreover, the experimenter can directly control the extent of any convexity of the effort cost function. Second, a choice-effort task allows researchers to capture effort directly rather than rely on a noisy proxy of effort (task performance). Finally, a choice-effort task is easier to create, and can be more efficiently administered. For example, unlike real-effort tasks, a choice-effort task does not need pre-testing, reducing researchers' costs in terms of time and money.

That being said, using a real-effort task can be advantageous for at least three reasons. First, a real-effort task is necessary to properly test some theories. For example, Farrell, Goh, and White (2014) find incentive pay increases the use of System 2 (deliberate) processing relative to System 1 (emotional) processing. Since a choice-effort task explicitly quantifies effort levels and the corresponding personal costs, such a task likely induces more System 2 processing than does a real-effort task. Thus, if researchers are testing theory involving more emotional, System 1 processing, then using a real-effort task may be more appropriate. Another example is the literature on the relative motivational effects of cash versus tangible rewards. Using mental accounting theory, studies in this literature stream test the prediction that tangible rewards will motivate greater effort than do cash rewards because individuals will categorize tangible rewards into mental accounts that are more novel, leading individuals to assess greater subjective value to tangible rewards (Presslee, Vance, and Webb 2013). However, a choice-effort task may make it difficult to test this prediction, because a choice-effort task would force participants to convert the tangible reward into monetary terms (e.g., the reward's market or retail value) and induce a joint evaluation that inhibits individuals' mental accounting processes (Shaffer and Arkes 2009).

Second, a real-effort task allows researchers to capture different dimensions of effort. Unlike a choice-effort task, which captures only one "type" of effort, a real-effort task can be used to capture effort duration (how long participants work), effort intensity (how hard participants work during a set amount of time), effort direction (what tasks participants choose to do), and strategy development (learning). Thus, a real-effort task is better suited for testing theories regarding different dimensions of effort rather than effort as a general construct.

Finally, using a real-effort task may lead to more generalizable theory than does a choice-effort task (Dutcher et al. 2015). Some scholars argue the abstract and artificial nature of choice-effort tasks limits the role of "real" effects (e.g., boredom and excitement), and such limits hinder the generalizability of (supported) theory (van Dijk, Sonnemans, and van Winden 2001; Charness and Kuhn 2011). For example, research finds individuals may derive utility from the act of working itself, which can moderate the effects of incentive pay on effort (Sprinkle 2000; Webb et al. 2013). Likewise, Dutcher (2012) finds participants' performance on a real-effort task depends on whether the task involves creativity, an insight that would not have been discovered using a choice-effort task. Thus, real-effort tasks allow researchers to capture other important phenomena related to effort, increasing the external validity of observed findings. In addition, a real-effort task allows researchers to provide insights into both the relationship between compensation schemes and effort, and the relationship between effort and performance. Such insights are critical to understanding the merits of different compensation schemes, as in the case of using incentive pay to motivate more creative output (Kachelmeier et al. 2008). In contrast, studies using a choice-effort task typically hold constant the relationship between effort and performance in order to isolate the relationship between compensation schemes and effort.

Importantly, the goal of our study is not to investigate the relative merits of using a real-effort versus a choice-effort task. Rather, we seek to highlight important considerations when designing a real-effort task, *given that such a task will be used to study effort*.

Appendix B - Summary of Accounting Studies using Real-Effort Tasks

Journal	Author(s)	Year	Task	Task Involves Effort Intensity?	Task Involves Multiple Dimensions of Effort?
TAR	Fisher, Frederickson, and Peffer	2000	Decode	Yes	No
TAR	Sprinkle	2000	Production Choice	No	Yes
AOS	Fisher, Frederickson, and Peffer	2002	Decode	Yes	No
TAR	Fisher, Maines, Peffer, and Sprinkle	2002	Decode	Yes	No
JMAR	Fessler	2003	Water-Jar Problems	Yes	No
JMAR	Fisher, Peffer, and Sprinkle	2003	Decode	No	No
TAR	Fisher, Maines, Peffer, and Sprinkle	2005	Decode	Yes	No
TAR	Magro	2005	Tax Scenario Analysis	Yes	Yes
AOS	Fisher, Frederickson, and Peffer	2006	Decode	Yes	No
AOS	Sprinkle, Williamson, and Upton	2008	Letter Search	Yes	No
CAR	Farrell, Kadous, and Towry	2008	Sandwich	Yes	Yes
JAR	Kachelmeier, Reichert, and Williamson	2008	Rebus Puzzle	Yes	Yes
JMAR	Church, Libby, and Zhang	2008	Decode	Yes	No
JMAR	Guymon, Balakrishnan, and Tubbs	2008	Decode	Yes	No
TAR	Hannan, Krishnan, and Newman	2008	Production Choice	No	Yes
BRIA	Libby and Thorne	2009	Lego Assembly	Yes	No

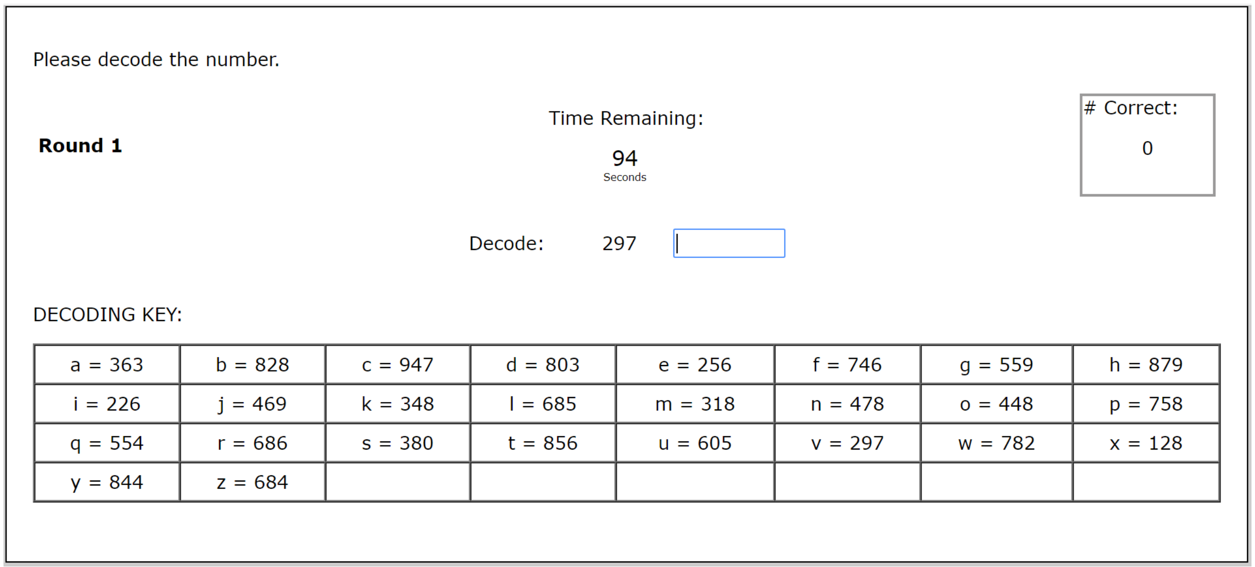
TAR	Kachelmeier and Williamson	2010	Rebus Puzzle	Yes	Yes
JMAR	Bailey and Fessler	2011	Jigsaw Puzzle	Yes	No
JMAR	Brüggen	2011	Decode	Yes	No
CAR	Christ, Emett, Summers, and Wood	2012	Data Entry	Yes	Yes
CAR	Hecht, Tafkov, and Towry	2012	Production Choice	Yes	Yes
JMAR	Farrell, Kadous, and Towry	2012	Sandwich	Yes	Yes
TAR	Chen, Williamson, and Zhou	2012	Idea Generation	Yes	No
BRIA	Naranjo-Gil, Cuevas-Rodríguez, López-Cabrales, and Sánchez	2012	Chip Assembly	Yes	No
TAR	Hannan, McPhee, Newman, and Tafkov	2013	Multiplication Problems, Anagrams	Yes	Yes
TAR	Tafkov	2013	Multiplication Problems	Yes	Yes
TAR	Webb, Williamson, and Zhang	2013	Letter Search	Yes	Yes
AOS	Newman and Tafkov	2014	Production Choice	No	Yes
AOS	Arnold and Gillenkirch	2015	Decode	Yes	No
JMAR	Kelly, Webb, and Vance	2015	Decode	Yes	No
TAR	Arnold	2015	Decode	Yes	No
TAR	Hales, Wang, and Williamson	2015	Sudoku Puzzles	No	No
AOS	Christ, Emett, Tayler, and Wood	2016	Data Entry	Yes	Yes
AOS	Thornock	2016	Maze Search	Yes	Yes
CAR	Kachelmeier, Thornock, and Williamson	2016	Letter Search	Yes	No
TAR	Choi, Newman, and Tafkov	2016	Sandwich	Yes	Yes
AOS	Kelly and Presslee	2017	Decode	Yes	No
CAR	Wang	2017	Letter Search	Yes	No
JMAR	Holderness, Olsen, and Thornock	2017	Decode	Yes	No

TAR	Farrell, Grenier, and Leiby	2017	Sandwich	Yes	Yes
AOS	Berger, Libby, and Webb	2018	Decode	Yes	No
JMAR	Kerstin, Merley, and Mellon	2018	Decode	Yes	No
TAR	Brüggen, Feichter, and Williamson	2018	Decode, Idea Generation	Yes	Yes
TAR	Chan	2018	Slider	Yes	Yes
TAR	Loftus and Tanlu	2018	Scrabble	Yes	Yes

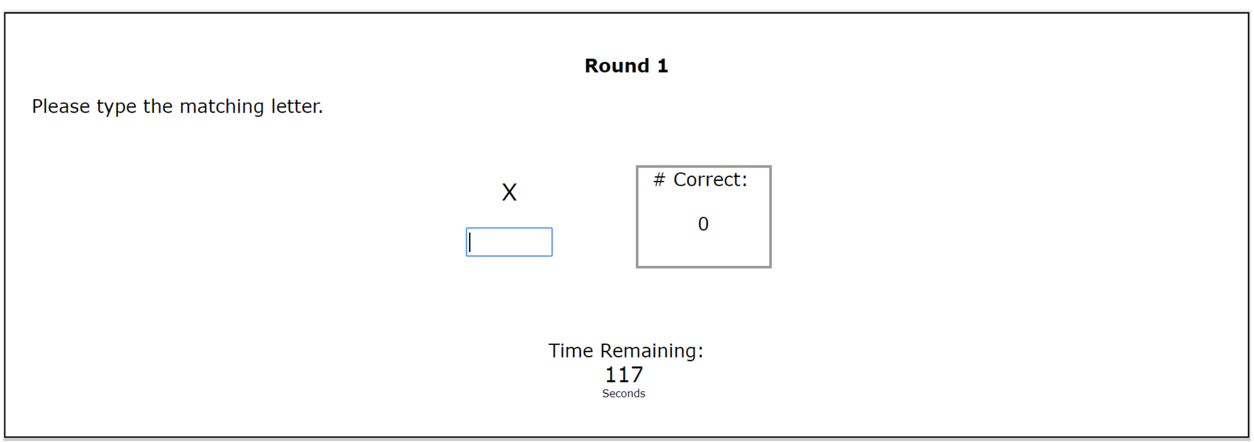
We generated this list of real-effort tasks by reviewing published and forthcoming studies from 2000-2018 in eight journals: TAR, JAR, JAE, CAR, AOS, RAST, JMAR, and BRIA.

Appendix C - Screenshots of Decode, Letter Search, and Slider Tasks

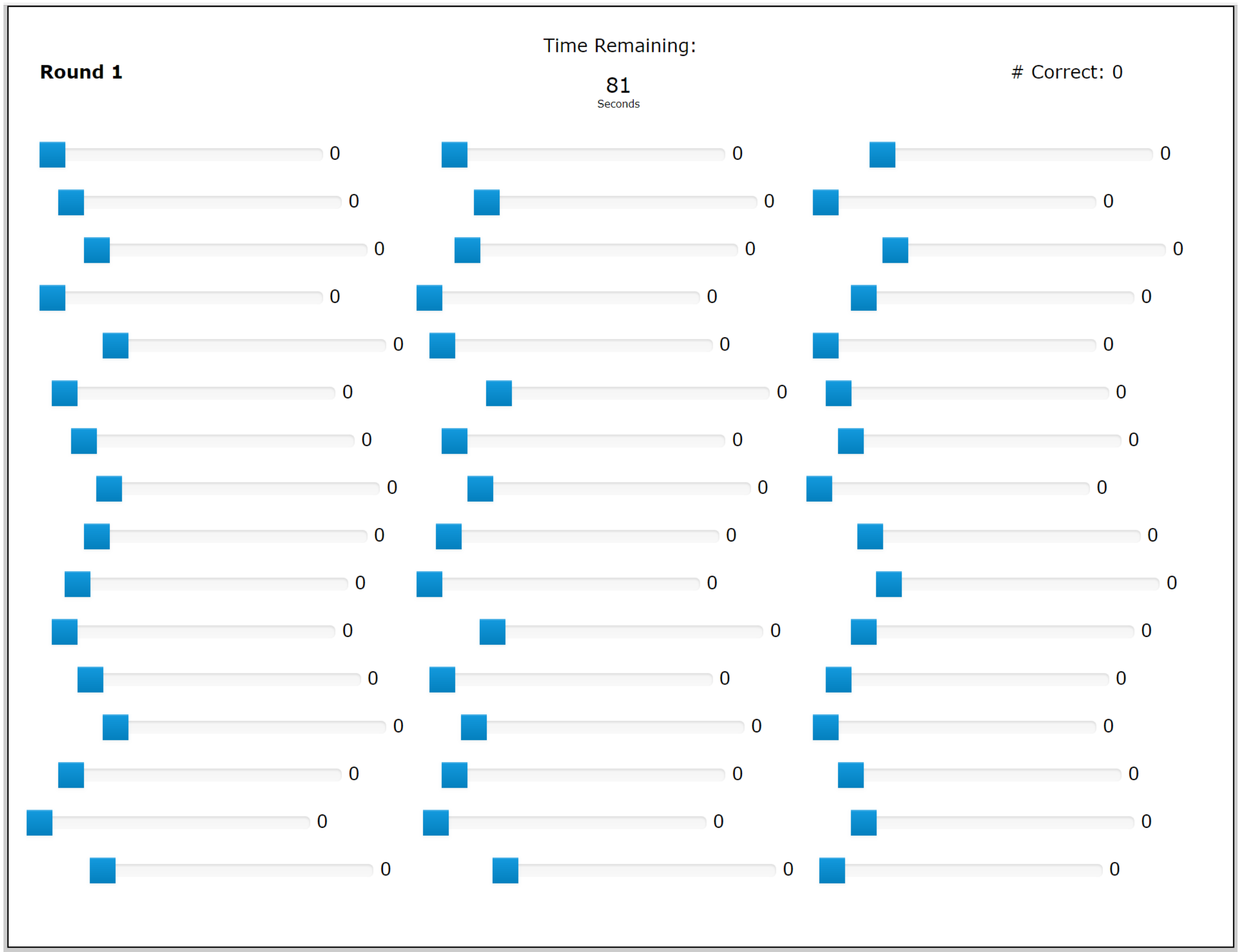
Decode Task



Letter Search Task



Slider Task



Appendix D - Risk Aversion[a]

Please make a selection for **each** of the 10 choices below. In each case, the choice is between two **hypothetical** alternatives that are labeled Alternative A and Alternative B. Alternative A always gives you a chance of winning a prize of $40 or $32, while Alternative B always gives you a chance of winning $77 or $2. However, each choice is different because the probability of your winning each amount changes.

	Alternative A You have:	Alternative B You have:	Your Choice Please choose Alternative A or Alternative B for each choice
1)	10% probability of winning $40 *and* 90% probability of winning $32	10% probability of winning $77 *and* 90% probability of winning $2	Alternative A Alternative B
2)	20% probability of winning $40 *and* 80% probability of winning $32	20% probability of winning $77 *and* 80% probability of winning $2	Alternative A Alternative B
3)	30% probability of winning $40 *and* 70% probability of winning $32	30% probability of winning $77 *and* 70% probability of winning $2	Alternative A Alternative B
4)	40% probability of winning $40 *and* 60% probability of winning $32	40% probability of winning $77 *and* 60% probability of winning $2	Alternative A Alternative B
5)	50% probability of winning $40 *and* 50% probability of winning $32	50% probability of winning $77 *and* 50% probability of winning $2	Alternative A Alternative B
6)	60% probability of winning $40 *and* 40% probability of winning $32	60% probability of winning $77 *and* 40% probability of winning $2	Alternative A Alternative B
7)	70% probability of winning $40 *and* 30% probability of winning $32	70% probability of winning $77 *and* 30% probability of winning $2	Alternative A Alternative B
8)	80% probability of winning $40 *and* 20% probability of winning $32	80% probability of winning $77 *and* 20% probability of winning $2	Alternative A Alternative B
9)	90% probability of winning $40 *and* 10% probability of winning $32	90% probability of winning $77 *and* 10% probability of winning $2	Alternative A Alternative B
10)	100% probability of winning $40 *and* 0% probability of winning $32	100% probability of winning $77 *and* 0% probability of winning $2	Alternative A Alternative B

[a]This measure of risk aversion was adapted from Holt and Laury (2002). I provide participants with ten scenarios and examine how they trade off safe alternatives with risky alternatives. In each scenario, participants choose between the safe alternative (probability p of winning $40 and probability 1-p of winning $32) and the risky alternative (probability p of winning $77 and probability 1-p of winning $2). In the first scenario, p = 0.10 and increases by 0.10 for each subsequent scenario. I measure risk aversion as the number of times the safe alternative is chosen before participants switched over to the risky alternative, so a higher number indicates a participant is more risk averse.

Bibliography

Armor, D. A., and S. E. Taylor. 1998. Situated optimism: Specific outcome expectancies and self-regulation. *Advances in Experimental Social Psychology* 30: 309-379.

Arnold, M. C. 2015. The effect of superiors' exogenous constraints on budget negotiations. *The Accounting Review* 90 (1): 31-57.

Arnold, M. C., and R. M. Gillenkirch. 2015. Using negotiated budgets for planning and performance evaluation: An experimental study. *Accounting, Organizations and Society* 43: 1-16.

Bailey, C. D., and N. J. Fessler. 2011. The moderating effects of task complexity and task attractiveness on the impact of monetary incentives in repeated tasks. *Journal of Management Accounting Research* 23: 189-210.

Baker, G. P., M. C. Jensen, and K. J. Murphy. 1988. Compensation and incentives: Practice vs. theory. *The Journal of Finance* 43 (3): 593-616.

Baron, R. M., and D. A. Kenny. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology* 51 (6): 1173-1182.

Berger, L., T. Libby, and A. Webb. 2018. The effects of tournament horizon and the percentage of winners on social comparison and performance in multi-period competitions. *Accounting, Organizations and Society* 64: 1-16.

Blasi, J., D. Kruse, and J. Sesil. 2001. Public Companies with Broad-Based Stock Options: Corporate Performance from 1992-1997. *The National Center for Employee Ownership.*

Blasi, J., D. Kruse, and J. Sesil. 2003. Sharing Ownership via Employee Stock Ownership. In *Ownership and Governance of Enterprises*, *Studies in Development Economics and Policy*, edited by L. Sun, London: Palgrave Macmillan.

Bonner, S. E., R. Hastie, G. B. Sprinkle, and S. M. Young. 2000. A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. *Journal of Management Accounting Research* 12: 19-64.

Bonner, S. E., and G. B. Sprinkle. 2002. The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research. *Accounting, Organizations and Society* 27: 303-345.

Brüggen, A. 2011. Ability, career concerns, and financial incentives in a multi-task setting. *Journal of Management Accounting Research* 23: 211-229.

Brüggen, A., and M. Strobel. 2007. Real effort versus chosen effort in experiments. *Economic Letters* 96: 232-236.

Brüggen, A., C. Feichter, and M. G. Williamson. 2018. The effect of input and output targets for routine tasks on creative task performance. *The Accounting Review* 93 (1): 29-43.

Chan, E. W. 2017. Promotion, relative performance information, and the Peter Principle. *The Accounting Review* 93 (3): 83-103.

Charness, G., and P. Kuhn. 2011. Lab labor: What can labor economists learn from the lab? In *Handbook of Labor Economics, Volume 4A*, edited by O. Ashenfelter and D. Card, 229-330. San Diego, CA: North Holland.

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

Choi, J., J. D. Clark, and A. Presslee. 2019. Guidance for Designing Real-Effort Tasks to Test for Motivational Effects. *Working paper.*

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

Christ, M. H., S. A. Emett, S. L. Summers, and D. A. Wood. 2012. The effects of preventive and detective controls on employee performance and motivation. *Contemporary Accounting Research* 29 (2): 432-452.

Christ, M. H., s. A. Emett, W. B. Tayler, and D. A. Wood. 2016. Compensation or feedback: Motivating performance in multidimensional tasks. *Accounting, Organizations and Society* 50: 27-40.

Church, B. K., T. Libby, and P. Zhang. 2008. Contracting frame and individual behavior: Experimental evidence. *Journal of Management Accounting Research* 20: 153-168.

Core, J. E., and W. R. Guay. 2001. Stock option plans for non-executive employees. *Journal of Financial Economics* 61: 253-287.

Core, J. E., W. R. Guay, and D. F. Larcker. 2003. Executive equity compensation and incentives: A survey. *FRBNY Economic Policy Review* April: 27-50.

Datar, S. M., and M. V. Rajan. 2014. *Managerial Accounting: Making Decisions and Motivating Performance*. Upper Saddle River, NJ: Pearson Education Inc.

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

Demski, J. S., and G. Feltham. 1978. Economic incentives in budgetary control systems. *The Accounting Review* 53 (2): 336-359.

Dutcher, E. G. 2012. The effects of telecommuting on productivity: An experimental examination. The role of dull and creative tasks. *Journal of Economic Behavior & Organization* 84: 355-363.

Dutcher, E. G., T. C. Salmon, and K. J. Saral. 2015. Is "real" effort more real? Working paper, Ohio University, Southern Methodist University, and Webster University Geneva.

Eppert, Anthony J. 2017. Trends in Designing Performance-Based Equity Awards. Andrews Kurth Kenyon LLP. Executive Compensation Webinar Series. https://www.huntonak.com/images/content/5/2/v2/52744/Trends-in-Designing-Performance-Based-Equity-Awards-081017.pdf.

Farrell, A. M., J. H. Grenier, and J. Leiby. 2017. Scoundrels or stars? Theory and evidence on the quality of workers in online labor markets. *The Accounting Review* 92 (1): 93-114.

Farrell, A. M., J. O. Goh, and B. J. White. 2014. The effect of performance-based incentive contracts on System 1 and System 2 processing in affective decision contexts: fMRI and behavioral evidence. *The Accounting Review* 89 (6): 1979-2010.

Farrell, A. M., K. Kadous, and K. L. Towry. 2008. Contracting on contemporaneous versus forward-looking measures: An experimental investigation. *Contemporary Accounting Research* 25 (3): 773-802.

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

Fessler, N. J. 2003. Experimental evidence on the links among monetary incentives, task attractiveness, and task performance. *Journal of Management Accounting Research* 15: 161-176.

Financial Engines. 2018. Market Summary: April 2018. *Financial Engines Education Center*. http://financialengines.com/education-center/market-summary-april-2018/.

Fisher, J. G., J. R. Frederickson, and S. A. Peffer. 2000. Budgeting: An experimental investigation of the effects of negotiation. *The Accounting Review* 75 (1): 93-114.

Fisher, J. G., J. R. Frederickson, and S. A. Peffer. 2002. The effect of information asymmetry on negotiated budgets: An empirical investigation. *Accounting, Organizations and Society* 27: 27-43.

Fisher, J. G., J. R. Frederickson, and S. A. Peffer. 2006. Budget negotiations in multi-period settings. *Accounting, Organizations and Society* 31: 511-528.

Fisher, J. G., L. A. Maines, S. A. Peffer, and G. B. Sprinkle. 2002. Using budgets for performance evaluation: Effects of resource allocation and horizontal information asymmetry on budget proposals, budget slack, and performance. *The Accounting Review* 77 (4): 847-865.

Fisher, J. G., L. A. Maines, S. A. Peffer, and G. B. Sprinkle. 2005. An experimental investigation of employer discretion in employee performance evaluation and compensation. *The Accounting Review* 80 (2): 563-583.

Fisher, J. G., S. A. Peffer, and G. B. Sprinkle 2003. Budget-based contracts, budget levels, and group performance. *Journal of Management Accounting Research* 15: 51-74.

Gill, D., and V. Prowse. 2012. A structural analysis of disappointment aversion in a real effort competition. *American Economic Review* 102 (1): 469-503.

Gill, D., and V. Prowse. 2015. A novel computerized real effort task based on sliders. Working paper, University of Oxford and Cornell University.

Goetzmann, W. N., and A. Kumar. 2008. Equity portfolio diversification. *Review of Finance* 12: 433-463.

Guymon, R. N., R. Balakrishnan, and R. M. Tubbs. 2008. The effect of task interdependence and type of incentive contract on group performance. *Journal of Management Accounting Research* 20: 1-18.

Hales, J., L. W. Wang, and M G. Williamson. 2015. Selection benefits of stock-based compensation for the rank-and-file. *The Accounting Review* 90 (4): 1497-1516.

Hall, B. J., and K. J. Murphy. 2003. The trouble with stock options. *Journal of Economic Perspectives* 17 (3): 49-70.

Hannan, R. L., V. B. Hoffman, and D. V. Moser. 2005. Bonus versus penalty: Does contract frame affect employee effort? In *Experimental Business Research*, edited by A. Rappaport and R. Zwick, Vol. II, Warren, MI: Springer Science and Business Media.

Hannan, R. L., G. P. McPhee, A. H. Newman, and I. D. Tafkov. 2013. The effect of relative performance information on performance and effort allocation in a multi-task environment. *The Accounting Review* 88 (2): 553-575.

Hannan, R. L., R. Krishnan, and A. H. Newman. 2008. The effects of disseminating relative performance feedback in tournament and individual performance compensation plans. *The Accounting Review* 83 (4): 893-913.

Hecht, G. W., I D. Tafkov, and K. L. Towry. 2012. Performance spillover in a multitask environment. *Contemporary Accounting Research* 29 (2): 563-589.

Herschung, F., M. D. Mahlendorf, and J. Weber. 2017. Mapping quantitative management accounting research 2002-2012. *Journal of Management Accounting Research* (forthcoming).

Hochberg, Y. V., and L. Lindsey. 2010. Incentives, targeting, and firm performance: An analysis of non-executive stock options. *The Review of Financial Studies* 23 (11): 4148-4186.

Hodge, F. D., S. Rajgopal, and T. Shevlin. 2009. Do managers value stock options and restricted stock consistent with economic theory? *Contemporary Accounting Research* 26 (3): 899-932.

Holderness, Jr., D. K., K. J. Olsen, and T. A. Thornock. 2017. Who are *you* to tell me *that*?! The moderating effect of performance feedback source and psychological entitlement on individual performance. *Journal of Management Accounting Research* 29 (2): 33-46.

Holt, C. A., and S. K. Laury. 2002. Risk aversion and incentive effects. *The American Economic Review* 95 (5): 1644-1655.

Hossain, T., and K. K. Li. 2014. Crowding Out in the Labor Market: A Prosocial Setting Is Necessary. *Management Science* 60 (5): 1148-1160.

Huck, S., G. Kirchsteiger, and J. Oechssler. 2005. Learning to like what you have – explaining the endowment effect. *Economic Journal* 115 (505): 689-702.

Ittner, C. D., R. A. Lambert, and D. F. Larcker. 2003. The structure and performance consequences of equity grants to employees of new economy firms. *Journal of Accounting and Economics* 34: 89-127.

Johnson, E. J., G. Häubl, and A. Keinan. 2007. Aspects of endowment: A query theory of value construction. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition* 33 (3): 461-474.

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 (5): 1669-1691.

Kachelmeier, S. J., B. E. Reichert, and M. G. Williamson. 2008. Measuring and motivating quantity, creativity, or both. *Journal of Accounting Research* 46 (2): 341-373.

Kachelmeier, S. J., T. A. Thornock, and M. G. Williamson. 2016. Communicated values as informal controls: Promoting quality while undermining productivity. *Contemporary Accounting Research* 33 (4): 1411-1434.

Kahneman, D., J. L. Knetsch, and R. H. Thaler. 1990. Experimental tests of the endowment effect and the coase theorem. *Journal of Political Economy* 98 (6): 1325-1348.

Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47 (2): 263-292.

Kahneman, D., and A. Tversky. 1984. Choices, values, and frames. *American Psychologist* 39: 341-350.

Kelly, K., and A. Presslee. 2017. Tournament group identity and performance: The moderating effect of winner proportion. *Accounting, Organizations and Society* 56: 21-34.

Kelly, K., R. A. Webb, and T. Vance. 2015. The interactive effects of *ex post* goal adjustment and goal difficulty on performance. *Journal of Management Accounting Research* 27 (1): 1-25.

Kersting, L., R. Marley, and M. J. Mellon. 2018. Tournament horizon: A marathon or a sprint? It depends upon the level of heterogeneity in ability among employees. *Journal of Management Accounting Research* (forthcoming).

Kim, E. H., and P. Ouimet. 2014. Broad-based employee stock ownership: motives and outcomes. *The Journal of Finance* 69 (3): 1273-1319.

Klein, K. J., and S. W. J. Kozlowski. 2000. *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. San Francisco, CA: Jossey-Bass.

Kluemper, D. H., L. M. Little, and T. DeGroot. 2009. State or trait: Effects of state optimism on job-related outcomes. *Journal of Organizational Behavior* 30: 209-231.

Kohn, A. 1993. *Punished by Rewards.* Boston, MA: Houghton Mifflin.

Kuang, X., and D. V. Moser. 2009. Reciprocity and the effectiveness of optimal agency contract. *The Accounting Review* 84 (5): 1671–94.

Lazear, E. P. 1991. Labor economics and the psychology of organizations. *Journal of Economic Perspectives* 5 (2): 89-110.

Libby, R., and J. Luft. 1993. Determinants of judgment performance in accounting settings: Ability, knowledge, motivation, and environment. *Accounting, Organizations and Society* 18 (5): 425-450.

Libby, T., and L. Thorne. 2009. The influence of incentive structure on group performance in assembly lines and teams. *Behavioral Research in Accounting* 21 (2): 57-72.

Libby, R., R. Bloomfield, and M. W. Nelson. 2002. Experimental research in financial accounting. *Accounting, Organizations and Society* 27: 775-810.

Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47: 13-37.

Locke, E. A., and G. P. Latham. 1990. *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice-Hall.

Loftus, S., and L. J. Tanlu. 2018. Because of "because": Examining the use of causal language in relative performance feedback. *The Accounting Review* 93 (2): 277-297.

Lorist, M. M., E. Bezdan, M. ten Caat, M. M. Span, J. B. T. M. Roerdink, and N. M. Maurits. 2009. The influence of mental fatigue and motivation on neural network dynamics; an EEG coherence study. *Brain Research* 1270: 95-106.

Luft, J. 1994. Bonus and penalty incentives contract choice by employees. *Journal of Accounting and Economics* 18: 181-206.

Luft, J., and M. D. Shields. 2003. Mapping management accounting: graphics and guidelines for theory-consistent empirical research. *Accounting, Organizations and Society* 28: 169-249.

Luke, D. 2004. *Multilevel modeling*. New Delhi, India: Sage Publication.

Magro, A. M. 2005. Knowledge, adaptivity, and performance in tax research. *The Accounting Review* 80 (2): 703-722.

Merchant, K. A., and W. A. Van der Stede. 2003. *Management Control Systems: Performance Measurement, Evaluation, and Incentives*. Essex, U.K.: Pearson Education Limited.

Möckel, T., C. Beste, and E. Wascher. 2015. The Effects of Time on Task in Response Selection – An ERP Study of Mental Fatigue. *Scientific Reports* 5 (10113): 1-9.

Morewedge, C. K., and C. E. Giblin. 2015. Explanations of the endowment effect: an integrative review. *Trends in Cognitive Sciences* 19 (6): 339-348.

Naranjo-Gil, D., G. Cuevas-Rodríguez, A. López-Cabrales, and J. M. Sánchez. 2012. The Effects of Incentive System and Cognitive Orientation on Teams' Performance. *Behavioral Research in Accounting* 24 (2): 177-191.

National Center for Employee Ownership (NCEO). 2014. Survey Data on Broad-Based Employee Ownership in the US. *A Statistical Profile of Employee Ownership*. www.nceo.org/articles/statistical-profile-employee-ownership.

Newman, A. H., and I. D. Tafkov. 2014. Relative performance information in tournaments with different prize structures. *Accounting, Organizations and Society* 39: 348-361.

Nonis, S. A., and D. Wright. 2003. Moderating effects of achievement striving and situational optimism on the relationship between ability and performance outcomes of college students. *Research in Higher Education* 44 (3): 327-346.

Oyer, P. 2004. Why do firms use incentives that have no incentive effects? *The Journal of Finance* 59 (4): 1619-1649.

Oyer, P., and S. Schaefer. 2005. Why do some firms give stock options to all employees? An empirical examination of alternative theories? *Journal of Financial Economics* 76: 99-133.

Pink, D. H. 2009. *Drive: The Surprising Truth About What Motivates Us*. New York: Riverhead Books.

Petersen, M. A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies* 22 (1): 435-480.

Presslee, A., T. W. Vance, and R. A. Webb. 2013. The effects of reward type on employee goal setting, goal commitment, and performance. *The Accounting Review* 88 (5): 1805-1831.

Roese, N. J. 1997. Counterfactual thinking. *Psychological Bulletin* 121 (1): 133-148.

Preacher, K. J., and A. F. Hayes. 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods* 40 (3): 879-891.

Scheier, M. F., and C. S. Carver. 1985. Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology* 4 (3): 219-247.

Scheier, M. F., C. S. Carver, and M. W. Bridges. 1994. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology* 67 (6): 1063-1078.

Segerstrom, S. C., S. E. Taylor, M. E. Kemeny, and J. L. Fahey. 1998. Optimism is associated with mood, coping, and immune change in response to stress. *Journal of Personality and Social Psychology* 74 (6): 1646-1655.

Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance* 19 (3): 425-442.

Shaw, J. D., and N. Gupta. 2015. Let the evidence speak again! Financial incentives are more effective than we thought. *Human Resource Management Journal* 25 (3): 281-293.

Shiller, R. J. 2001. *Irrational Exuberance*. Princeton, NJ: Princeton University Press.

Shiller, R. J. 2016. *Irrational Exuberance*, 3rd ed. Princeton, NJ: Princeton University Press.

Shiller, R. J. 2018. Shiller PE Ratio. www.multpl.com/shiller-pe/.

Shu, S. B., and J. Peck. 2011. Psychological ownership and affective reaction: Emotional attachment process variables and the endowment effect. *Journal of Consumer Psychology* 21 (4): 439-452.

Sobel, M. 1982. Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, edited by Leinhardt, S., 144-158. Washington, DC: American Sociological Association.

Sprinkle, G. B. 2000. The effect of incentive contracts on learning and performance. *The Accounting Review* 75 (3): 299-326.

Sprinkle, G. B., M. G. Williamson, and D. R. Upton. 2008. The effort and risk-taking effects of budget-based contracts. *Accounting, Organizations and Society* 33: 436-452.

Tafkov, I. D. 2013. Private and public relative performance information under different compensation contracts. *The Accounting Review* 88 (1): 327-350.

Taylor, S. E., M. E. Kemeny, L. G. Aspinwall, S. G. Schneider, R. Rodriguez, and M. Herbert. 1992. Optimism, coping, psychological distress, and high-risk sexual behavior among men at risk for acquired immunodeficiency syndrome (AIDS). *Journal of Personality and Social Psychology* 63 (3): 460-473.

Thaler, R. 1980. Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization* 1: 39-60.

Thornock, T. A. 2016. How the timing of performance feedback impacts individual performance. *Accounting, Organizations and Society* 55: 1-11.

van Dijk, F., J. Sonnemans, and F. van Winden. 2001. Incentive systems in a real effort experiment. *European Economic Review* 45: 187-214.

Vroom, V. 1964. *Work and Motivation.* New York, NY: John Wiley.

Wang, L. W. 2017. Recognizing the best: The productive and counterproductive effects of relative performance recognition. *Contemporary Accounting Research* 34 (2): 966-990.

Weaver, R., and S. Frederick. 2012. A reference price theory of the endowment effect. *Journal of Marketing Research* 49: 696-707.

Webb, R. A., M. G. Williamson, and Y. Zhang. 2013. Productivity-target difficulty, target-based pay, and outside-the-box thinking. *The Accounting Review* 88 (4): 1433-1457.

Weinstein, N. D. 1980. Unrealistic optimism about future life events. *Journal of Personality and Social Psychology* 39 (5): 806-820.

WorldatWork and Deloitte Consulting LLP. 2014. Incentive pay practices survey: Publicly traded companies. https://www.worldatwork.org/docs/research-and-surveys/survey-brief-incentive-pay-practices-survey-publicly-traded-companies.pdf.