An Experimental Test of Gaming Incentives

Sheheryar Banuri, Katarína Danková, and Philip Keefer[1]

May 2024

DRAFT: Please do not cite without permission

Abstract

In recent years, many organizations have undertaken efforts to introduce gaming incentives in order to enhance employee effort and user engagement. However, these incentives are not well understood, and subsequently there is a dearth of systematically controlled empirical evidence on their effects, as well as their interaction with other forms of motivation. We present the first comprehensive test of gaming incentives using a real effort lab experiment with a standard lab task adapted to allow the inclusion of gaming incentives. We test the effect of these incentives under two types of extrinsic incentives, an unconditional wage, and a piece rate. In addition, we also measure subject task motivation and explore interaction with intrinsic (task based) incentives. We report three main findings: first, gaming incentives increase effort when extrinsic incentives to exert effort are weak. When extrinsic incentives are strong, gaming incentives have no additional impact on effort, showcasing diminishing returns to multiple forms of incentives. Finally, some forms of gaming incentives are successful in increasing effort among the least task motivated, a property that is similar to extrinsic incentives. Implications for organizations seeking to implement gaming incentives are discussed.

Keywords: Intrinsic motivation, extrinsic motivation, performance, feedback, gamification, effort
JEL Classification Codes: J24, M12, C91

---

[1] Banuri: School of Economics, University of East Anglia, and Hughes Hall, University of Cambridge (e-mail: s.banuri@uea.ac.uk); Dankova: Central European Labour Studies Institute, Slovakia (e-mail: katarina.danko@gmail.com); Keefer: Inter-American Development Bank (e-mail: pkeefer@iadb.org). The authors have no relevant or material financial interests that relate to the research described in this paper. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. We wish to thank Maroš Servatka (MGSM Sydney), Oana Borcan (UEA Norwich) and participants of the 4th Lund Conference on Games, Interaction, Reasoning, Learning and Semantics 2016 and ESA European meeting 2016 in Bergen for valuable comments and suggestions. We would like to especially thank the associate editor and two anonymous referees for extremely helpful comments that have substantially improved the paper. Financial support was provided by the University of East Anglia. Any errors remain our own.

I. Introduction

A persistent question in education, marketing, management, and behavioral economics is how best to use performance feedback to increase effort. Most social science research in this area focuses on how to deliver feedback (absolute versus relative performance, for example) and the frequency of delivery. Few papers have considered making the feedback itself motivating (and fun). Using a real effort task common to the experimental literature, we add gaming incentives to the task and ask whether (i) they increase effort, (ii) whether they interact with pecuniary incentives, and (iii) whether they interact with task-based incentives.

A central principle of the gaming industry is that context of feedback matters: the way in which feedback is delivered affects effort by increasing motivational "affordances"[2] and gameful experiences (Hamari, Koivisto, and Sarsa, 2014; Detering et al. 2011).[3] We ask whether gaming incentives (otherwise known as gamification – the integration of game elements into standard performance feedback) motivates individuals to exert more effort. If so, which gaming incentives matter, and how do they interact with other types of incentives? We conduct a real-effort lab experiment with a baseline (no incentive) condition, four treatments each of which adds the four most common gaming incentives, and a fifth treatment with all four gaming incentives combined. In addition, we also conduct two additional treatments with a high-powered extrinsic incentive (piece rate), with and without gaming incentives to identify how these incentive systems interact. We also measure task motivation using a survey. Our results show that gaming incentives increase effort but exhibit diminishing returns: under low-powered extrinsic incentives, subjects provide significantly greater effort when some types of gaming

[2] An "Affordance" is a property of an environment that (when perceived) it provides actions to users that can satisfy certain needs (Gibson, 1979; Norman, 1999; Zhang, 2008). Motivational affordances are thus defined as perceived properties of an environment that can satisfy motivational needs for its user. Gaming elements increase motivational affordances by (once engaged with) providing sources of intrinsic motivation such as a sense of accomplishment or achievement.

[3] Motivation refers to a desire to exert effort, and comes from a variety of sources, including external controls, incentives, punishments, rewards, etc. (e.g., Herzberg, 1966; Porter & Lawler, 1968; Staw, 1977).

incentives are present. However, under high powered extrinsic incentives, effort does not increase (and even decreases slightly). Second, the increases in effort stem from the use of Leaderboards (which induce competition and is similar to relative performance feedback) and Stories/Themes (which induce meaning or purpose to the task). Finally, we measure task motivation to understand whether the effects of gaming incentives induce effort among the least motivated. We find that only one treatment, the use of stories, increases effort among the least motivated by the task. We discuss implications of these results for organizations seeking to implement gaming incentives for their workforce.

Game developers invent elaborate point systems, symbolic rewards, leaderboards and storylines, none of which change players' core tasks, to elicit interest in their games (see, for example, Deterding, et al. 2011). This interest is through the implementation of motivational affordances (Zhang, 2008; Deterding, 2011). Need based theories of motivation (such as Self Determination Theory) argue that individuals seek out experiences that satisfy motivational needs (Ryan and Deci, 2000; Deterding, 2011). Games fulfil these motivational needs by allowing individuals to feel competent or autonomous in their own environment. Many games ask individuals to engage in the same action (rolling a die or manipulating a controller), but manipulate the feedback from the action to different outcomes, which themselves can be motivating. These gaming elements obviously have an impact: not only do people exert substantial effort in games, whether measured in time or exertion; they also spend billions of dollars for the right to play them. For example, according to Newzoo's (a leader in games, e-sports and mobile intelligence) Global Games market report[4] an estimated the global gaming market to be worth 184 billion USD in 2023, with an estimated 3.3 billion gamers. Furthermore,

---

[4] Accessible at: https://newzoo.com/resources/trend-reports/newzoo-global-games-market-report-2023-free-version

according to Nielsen Gamer Insights study,[5] the gaming industry is currently 2.5 times the size of the global box office and music revenues combined.

Gaming elements have already found their way into applications outside of the gaming industry. Marketers have introduced gaming elements in a variety of contexts to increase demand, for example for social networking services (Facebook), games themselves (Angry Birds) and location-based services (Foursquare) (Hamari, 2013). Gamification has already been applied in the promotion of greener energy consumption (EcoIsland), taking care of one's health (Fitocracy), running (Zombies, Run!), and even to tracking one's life aspirations (Mindbloom). Large private sector firms, such as Uber, have utilized gaming techniques to elicit effort from their employees: Uber drivers are given badges for achieving certain quality benchmarks, and utilize ratings systems to induce better interactions (both for drivers and customers).[6] This phenomenon raises a natural question in the context of a large body of research on non-pecuniary motivation and effort choices within organizations (e.g. Falk, Gächter, and Kovács, 1999; Fehr and Falk, 1999). Does the integration of game elements into organizational tasks – gamification – motivate individuals to work harder (i.e. exert more effort) on those tasks? And how do these game elements interact with existing incentive schemes?

Gamification manipulates performance feedback and presents it in a way that is itself motivating. Hamari, Koivisto, and Sarsa (2014) offer a useful review of the gamification literature and find that the three main types of gaming features used are points, badges, and leaderboards. In addition to this, levels, and a story/theme were amongst the next most common features. The use of points, badges and leaderboards is self-explanatory. The use of a story/theme is to provide context and motivation for the user to continually engage in effort.

---

[5] Accessible at: https://nielseniq.com/global/en/insights/report/2021/gaming-industry-growth-as-the-ultimate-entertainer/

[6] See more of Uber's use of gamification here: https://www.nytimes.com/interactive/2017/04/02/technology/uber-drivers-psychological-tricks.html

The point is to present the feedback in such a way as to indicate progress towards some well-defined end goal. The engagement from the story allows for an additional form of motivation that interacts with motivation arising from the task itself.[7]

Research on performance feedback argues that clear goals and frequent performance feedback, unrelated to compensation, improves performance. The feedback could come in the form of evaluations of individual performance relative to some absolute standard, as in the management literature (e.g. Bloom and Van Reenen, 2007), or relative to others (Eriksson, Poulsen and Villeval, 2009). Psychologists argue that the central element of feedback is that it focuses on individuals' effort on the task rather than on their ability in general (Kluger and DeNisi, 1996). The considerable literature in economics on feedback and performance generally finds that feedback has heterogenous effects across agents (Eriksson et al. 2009; Azmat and Iriberri, 2010; Lizzeri, Meyer, and Persico, 2002; Ederer, 2010; Bandiera, Barankay, and Rasul., 2007; Blanes-i-Vidal and Nossol, 2011; List and Rasul, 2011; Barankay, 2012; Dubey and Geanakoplos, 2005, 2010). However, the literature mainly considers the role of information about performance in the task itself, rather than on generating additional motivational factors using performance feedback.

Finally, an important literature within economics and management highlights the role of goal setting increasing performance (Locke and Latham, 1990; 2002; Suvorov and van de Ven, 2008; Hsiaw, 2013; Koch and Nafziger, 2011; Corgnet, Gómez-Miñambres and Hernán-Gonzalez, 2015). This literature points to non-binding goals improving performance, due to goals serving as reference points and allowing individuals to regulate their own behavior (Koch and Nafziger, 2011), or the behavior of their workers (Corgnet, Gómez-Miñambres and Hernán-

[7] See Sailer and Homner (2020) for a meta-analysis on gamification in learning environments, and Saleem, Noori, and Ozdamil (2022) for a literature review in an e-learning context. Klock et al. (2020) provide a useful review on tailoring gamification to individuals, and Krath and Korflesch (2021) provide an excellent review of the theory underpinning gamification.

Gonzalez, 2015). The introduction of goals sets reference points for workers which then interacts with their intrinsic motivation, and creates similar motivational affordances to provide individuals with a sense of achievement.

Gamifying tasks incorporates the goal setting and performance feedback attributes that are central to the feedback literature. Relative performance comparisons, such as leaderboards, are central to both games and feedback. Games go further, however, by introducing gaming elements that do not change the task, the goals, nor the performance information provided to players. Instead, they add other intangible elements that are also expected to increase performance by making goal achievement more interesting, motivating, or "fun" (i.e. satisfying additional intrinsic motivational needs, such as the need for autonomy, competence, and relatedness – Cassar and Meier, 2018; Ryan and Deci, 2000). Psychologists debate what precisely makes these gaming elements "fun" and engaging (Detering, 2011). A common thread running through the debate overlaps with the feedback literature: gaming elements are a sophisticated form of feedback: they provide accelerated feedback, clear goals and challenging tasks (Hamari 2013). However, recognizing the importance of context and narrative in games, others argue that gaming elements give "meaning" to a particular task (Deterding, 2011; McGonigal, 2011) and provide additional sources of motivation.

Despite the success of these elements in encouraging people to purchase and play games, little previous research explicitly examines the effects of gamification on effort. Hamari, Koivisto, and Sarsa (2014) identify 24 papers in their review of research on the effects of gamification. Substantively, most research focuses on the effects of gamification on learning in educational contexts. Methodologically, the research is largely descriptive (e.g., lacking control groups) and yields ambiguous evidence on the effects of gamification on effort. An important exception is research conducted by Mekler, et al. (2013). They allow effort to affect whether subjects accumulate points, shift their position on leaderboards, or achieve higher levels, and ask whether these gaming elements affect effort in a real effort task: image annotation. Relative to a no-feedback baseline, gamification increases performance. They do not address the interaction of gamification, extrinsic and task motivation, the focus of this paper.

Furthermore, no prior experimental literature has examined a central gaming element that we study here, a narrative that changes with effort. The gaming elements we examine relate to, but are distinct from, previous literature that has examined non-pecuniary motivations to engage in a task (e.g., Kosfeld and Neckermann, 2011; Ashraf, Bandiera, and Jack, 2014; Bradler et al. 2016; Ariely, Kamenica, and Prelec, 2008; Banuri, Keefer, and de Walque 2021): this work particularly examines the effects of task motivation, but also status and peer effects. For example, Kosfeld and Neckermann (2011) show that effort increases when top performers on a task are promised a congratulatory card signed by the managing director of the organization that contracted the task. This award conveys social recognition and status, precisely the motivations that the researchers seek to examine. Gaming incentives, in contrast, generally do not invoke pro-social missions; they carry no additional status (leaderboards, for example, reveal only a player's relative position but not the identity of other players), and they are separate from the task. However, they operate by satisfying individual needs for competence, autonomy, and/or relatedness, and hence are themselves intrinsically motivating.

We also consider an important issue raised by prior literature, whether the effects of additional motivating elements depend on the pre-existing extrinsic and intrinsic motivations of individuals (Frey and Oberholzer-Gee 1997; Deci and Ryan 1985). Extrinsic incentives reduce intrinsic motivation in certain contexts (Deci, Koestner, and Ryan, 1999; Frey and Oberholzer-Gee, 1997; Ryan and Deci, 2000; Deci and Ryan, 1985). For example, Ashraf, Bandiera, and Jack, (2014) find that agents who are offered non-financial rewards exert more effort than those offered financial margins or volunteer contracts. Banuri, Keefer and de Walque (2021) examine the interaction of two intrinsic motivations, pro-sociality and task enjoyment, and find that the first has no additional effect on effort when the second is high. Casas-Arce and Martinéz Jerez (2009) study the effects of contests when ability is heterogenous. They find that winning participants decrease their effort as their lead extends, whereas the effort of trailing participants fades only when the gap to the winning position is very large.

Our experiment allows us to comprehensively test whether gaming incentives increase effort under two extrinsic motivation conditions: a low-powered incentive (a wage unrelated to effort), and a high-powered incentive (a piece rate for effort), in line with prior research examining the interaction of different forms of motivation (see for example, Ariely, Bracha and Meier, 2009; Carpenter and Myers, 2010; Friedrichsen and Engelmann, 2017; and Banuri, Keefer, and de Walque, 2021). Eriksson et al. (2009) find that feedback on performance generates higher effort in tournament settings, but not in piece rate settings. We compare feedback with gaming elements across flat salary and piece rate settings.

Our experiment centers on a real-effort decoding task, translating numbers into words, very similar to ones commonly used in the literature (Laughlin, Bonner, and Miner, 2002; Laughlin et al. 2006; Erkal, Gangadharan, and Nikiforakis, 2011; Nikiforakis, Noussair, and Wilkening, 2012; Dato and Nieken, 2014; McDonald et al. 2013; Johnson and Ramalingam, 2016; Neitzel and Saaksvuori, 2013; among others). In the baseline treatment, subjects receive a wage (unrelated to effort) and are continually informed about their performance in terms of units (basic feedback, consisting of continuous feedback on the number of words decoded). We evaluate effort on this task under different gaming incentive schemes: points systems, symbolic rewards, relative performance, and stories/themes, the four pillars of "gamification": the transformation of tasks into "games" (Deterding, et al. 2011). In addition, because modes of extrinsic compensation vary substantially across organizations, we also vary extrinsic incentives, by substituting the flat salary with a high-powered extrinsic incentive (piece rate) to perform the task.

The four gaming incentives are the building blocks of many enhanced feedback mechanisms. Organizations often integrate these four design features into their feedback systems. Primary school children receive gold stars, in addition to grades; Uber drivers receive badges for achieving quality benchmarks; and employers regularly provide private information to employees about their ranking relative to other employees. In addition, some performance feedback schemes explicitly link individual performance to the broader mission of the organization. Schools may encourage performance by showing how grades in a class contribute to life success, firms may link individual workers' sales and cost savings to the broader position of the firm in the industry, and marketing campaigns attempt to link individual consumption of a product to the individual's position in the world.

Our paper makes three main contributions to the literature on effort and gaming incentives: First, we offer a comprehensive test of whether gaming incentives increase effort. We show that they do and provide evidence on which incentives induce effort. Second, we show that gaming incentives exhibit diminishing returns: they increase effort when extrinsic incentives are low, but do not increase effort when extrinsic incentives are high. Third, we conduct some exploratory analysis with task motivation and show how gaming incentives induce effort from the least task motivated.

To our knowledge, this study is the first to provide evidence for gaming incentives and their interaction with extrinsic and intrinsic incentives. Our results are important for the efficiency of operations in organizations, and for managers seeking to enhance motivation for work tasks.

II. Experimental design

The experiment entails a basic real effort task in which subjects receive five numbers and a code that subjects use to decode the numbers into words, one letter at a time (see Figure 1). The experiment uses 6 distinct code sets (each randomly generated) which assigns one number to each character of the alphabet. We use six different code sets to mitigate learning. Using a single code allows subjects to memorize the code and become faster as they continue the task. Resetting the code inhibits this. Each round of the effort task lasts for 2 minutes, during which subjects decode a preset list of 5-letter words from the given numbers. Researchers have frequently employed this effort task to generate endowments in lab experiments (see, for example: Erkal, Gangadharan, and Nikiforakis, 2011; Neitzel and Saaksvuori, 2013; among others). It is useful for its simplicity, and because it allows us to implement narratives into a simple task, something that is fairly challenging for other real-effort tasks.

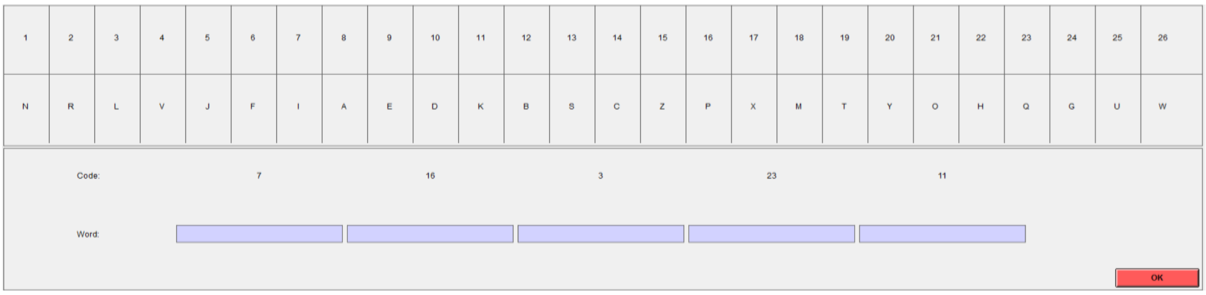


Figure 1: Decoding task

In the baseline, subjects receive an unconditional wage of 400 tokens (4 GBP) and are asked to undertake the task for at least one round. During each round, and at the end of the first round, subjects are given basic feedback on their performance. This type of feedback is standard across organizations and lacks any gaming incentives. During the round, subjects are provided with a running list of the words that they have decoded correctly, as well as the total number of words coded correctly, the percentage of words coded correctly, as well as the tokens earned. At the end of the round, subjects are provided with a history table containing the round number, the number of words coded correctly, and the percentage of words coded correctly in the round. The history table contains a history of the preceding three rounds. Subjects are then asked whether they would like to continue the task for an additional 2 minutes, or to end the task. As the instructions make clear, subjects understand that ending the task means starting the exit survey, after the completion of which they are free to leave. Subjects are not told how many times they can continue the task, but they can continue for a maximum of 13 rounds, after which the exit survey automatically begins.[8] Hence, in the baseline, subjects are paid an unconditional wage and can engage in the task for between 2 and 28 minutes (a minimum of 1 round, and a maximum of 14 rounds). The is our main dependent variable: the number of times subjects chose to continue the task (we also check for robustness to alternative measures of effort, tables provided in the appendix). This baseline offers low extrinsic incentives to exert effort.

We conduct a total of 7 treatments (in addition to the baseline), adding either gaming incentives, or extrinsic incentives. Table 1 presents the number of subjects in each treatment. The experiment randomized subjects to treatment within each session, hence the number of subjects is different across treatments due to randomization. We ran the experiment with the aim to reach a minimum of 40 subjects in each treatment. Based on the mean and standard deviation of our main dependent variable in the baseline (the number of times subjects continued in the task), and using standard assumptions (alpha = 0.05 and beta = 0.8), a sample of between 40 and 55 subjects per treatment means that we are powered to detect effect sizes of between 0.64 and 0.54 standard deviations, which is between a medium (0.5 standard deviations) and a large (0.8 standard deviations) effect size. While sample sizes of this magnitude are common in controlled physical laboratory environments, we note that we are not powered to detect smaller effects, and hence our results need to be interpreted with some caution.

Table 1: Experimental design

Treatment	Subjects
Baseline	43
Gaming feature 1 (Points)	41
Gaming feature 2 (Badges)	45
Gaming feature 3 (Leaderboard)	40
Gaming feature 4 (Story/Theme)	47
All Gaming features	51

[8] The decision not to inform the subjects of the maximum number of rounds reduces the impact of focality. Informing subjects of the end may have biased decisions upwards depending on their levels of reciprocity. To avoid this, we did not provide the subjects with a maximum number of rounds.

Piece rate	55
Piece rate + All gaming features	40

The first four treatments modify the baseline to include each of our four gaming incentives individually: Points, Badges, Leaderboard, and Story/Theme, in line with the most common gaming incentives reported in the literature on gamification (Hamari, Koivisto, and Sarsa, 2014).

Gaming incentive I – Points: Points systems provide a numerical score of task performance. By themselves, points convey the same information as the feedback used in the baseline, which informs subjects of the number of words coded correctly. However, one way to add a gaming incentive and increase motivational affordances is to include modifications to the rate of point accumulation. To do this, we award subjects 50 points for each word that they decode correctly and then increase the points by a multiplier that starts at 1 and increases by 0.1 for each consecutive word that subjects decode. If subjects make a mistake, the multiplier resets to 1. This feature differs from simple performance feedback used in the baseline, by adding a points multiplier which can be independently motivating for our subjects if they care about the multiplier in addition to their own points score.

Gaming incentive II –Badges: Gaming incentives can induce additional motivational affordances by offering symbolic rewards, such as badges, for goal achievement. These rewards are unrelated to the task, convey no additional information about goal-achievement, and have no extrinsic value. Nevertheless, they create a desire by individuals to accumulate these rewards, thereby inspiring greater effort. Badges (typically a graphic or a picture) are a common symbolic reward, particularly in games, but they can also be found in organizations. They provide recognition to the individual who has achieved some goal; different badges can be given for greater accomplishment.

For the Badges gaming incentive, we introduce a series of (on-screen) rewards based on preset targets. For any round, if subjects decode a minimum of 4 words they receive a bronze badge; with a minimum 7 words, a silver badge, and so on. The more words subjects decode (in a given round), the more valuable the material associated with the badge they receive (bronze, silver, gold, etc.). The badges do not reveal any information about the number of rounds available to the subjects. They are dependent solely on the number of words decoded within a single round. Table 2 displays the badges achievable by the subjects.

Table 2: Targets and corresponding badges

Number of words decoded correctly	Level	Medal
0-3	None	None
4-6	Bronze	
7-9	Silver	
10-12	Gold	
13-15	Diamond	
16-18	Platinum	
18+	Titanium	

Gaming incentive III – Leaderboard: Leaderboards provide relative performance feedback to induce competition. Individuals' positions on leaderboards can be public or private knowledge, but in both games and organizations, they are often private. For example, employees are typically told confidentially where their performance ranks compared to the performance of others in the organization, their ranking is also private information. Leaderboard rankings are also private knowledge here.

Subjects are assigned a rank and given a place on the leaderboard that corresponds to the number of words they decode correctly in a round. With performance within a round, subjects achieve higher ranks relative to their peers.

Typically, experiments that implement competition have two individuals simultaneously competing with each other. This turns out to be a challenge in our experiment because the key outcome variable is how long subjects engage in the task. In order to keep the implementation parallel to other treatments, we constructed a leaderboard from effort exerted by subjects in sessions conducted during the first day of the experiment.[9] Table 3 displays how the leaderboard was shown to the subjects. At the end of each round, subjects were assigned a rank on the leaderboard based on the number of words they decoded correctly within the round. Subjects were ranked on the leaderboard according to their highest score in the session.

Table 3: Example of a leaderboard

Round	Number of words decoded correctly	Percentage of words decoded correctly	Points scored	Rank on leaderboard
7	**7**	**100**	**455**	**RANK 62 OUT OF 75**

Gaming incentive IV – Story/theme: The story gaming incentive links subject effort to a larger purpose, and provides an additional source of motivation: Effort unveils a short story in each round. In games, a fundamental technique to encourage effort in (and demand for) a game is to create a narrative around otherwise mundane tasks (e.g. rolling dice or moving cursors)

[9] Since the experimental design has a large number of treatments, we randomly assigned treatments to sessions at the outset of the experiment. The first time a leaderboard was needed was on session 4, which was conducted on day 2 of the experiment. Hence, we were able to take data from sessions conducted on day 1 (with 75 subjects), and use them to construct a leaderboard, based on the maximum number of words decoded by each subject during their session. All subjects were informed that the data used to construct the leaderboard was from prior sessions (see instructions in table A.2 in appendix A).

generating a source of motivation outside the task itself. Ariely, Kamenica and Prelec (2008) show that actions that reduce the meaning or perceived value of tasks reduce effort.

We implement the story gaming incentive by developing 14 short stories, taken from a variety of sources, including fables from different cultures. We then edited the stories for similar lengths and maximized the number of 5-letter words contained within each story. These words were then eliminated from the story. Subjects engage in the same decoding task, but now each decoded word advances the story.[10]

Each round contains a different story. The amount of the story that is revealed depends on the number of words that subjects decode correctly. The more words they decode, the more of the story they get at the end of the round. Below is an example of a story used in the experiment (with the coded words underlined):

> *On a <u>rainy</u> day, a <u>young</u> girl <u>found</u> a <u>small</u> <u>black</u> <u>snake</u>. It spoke to her: "If you kiss me, I will transform into a prince." The girl took the <u>snake</u> and put it in her <u>purse</u>. <u>Again</u>, the <u>snake</u> <u>cried</u>, "I have a <u>power</u> to transform from a <u>snake</u> to a prince!" The girl did not <u>react</u>. The <u>snake</u> got <u>upset</u> and <u>cried</u> <u>aloud</u>, "What is <u>wrong</u> with you, don't you <u>fancy</u> <u>being</u> my <u>queen</u>?" The girl looked at the <u>snake</u> and said, "I am a <u>pilot</u>, I <u>spend</u> sixteen <u>hours</u> a day on a <u>plane</u>. I do not have time for a prince, but a talking <u>snake</u> is fantastic!"*

Subjects receive no additional reward for completing the story: their pecuniary compensation is unrelated to the outcome of the story. Nor does completing the story demand any different or additional effort from them. It is merely the case that if they happen to work harder on the underlying task, they advance the narrative.

Beyond the individual gaming incentive treatments, we conduct three additional treatments: one which combines all gaming incentives simultaneously, and tests for the joint

---

[10] Note that if a word is repeated in the story, it still only needs to be decoded a single time.

effect of these gaming incentives under an unconditional wage. The piece rate treatment implements a conditional wage: it asks participants to engage in the same effort task, but instead of providing subjects with an unconditional payment, subjects are informed that they will be paid a piece rate of 3.6 tokens per word coded correctly.[11] As with the salary treatment, subjects choose whether to continue or end the task at the end of each 2-minute round of decoding. The final treatment adds gaming incentives to the piece rate extrinsic incentive.

It is important to note that none of the gaming incentives matter across rounds. Each incentive pertains to the intensity of effort provided within each round. However, our core outcome measure is the number of times subjects repeated the task. This is theoretically independent of the gaming incentives offered because the incentives are not towards doing more rounds but improving effort within rounds. However, they motivate individuals to participate in additional rounds if they care about the gaming incentives at all.

To sum up, we report eight treatments in total. A baseline, four individual gaming incentives, simultaneous gaming incentives, a piece rate, and a piece rate with simultaneous gaming incentives. Our hypotheses are similarly straightforward: based on the literature (largely in computer science) on the positive effects of gamification, we expect positive impacts on effort for each individual gaming incentive, as well as the combined gaming incentive under low extrinsic incentives. Note that we also expect diminishing returns to additional sources of incentives (in line with the literature on motivation crowding, and as detailed out by our model in appendix C), meaning that effort does not increase under high-powered extrinsic incentives.

[11] The piece rate was selected to equalize earnings across treatments to minimize differences in behavior due to differences in income. The first two sessions were conducted with a flat salary of 400 tokens. Based on this, we found that subjects decoded an average of 7.91 words in the first round (which all subjects had to complete). Since we assumed subjects would continue the task for the maximum of 14 rounds, the piece rate was computed by taking the flat salary (400), dividing by the number of rounds (assumed to be 14) and the average output per round (based on the first two sessions: 7.91), and rounded to the nearest decimal to yield 3.6 tokens per word decoded. On average, subjects earned 414.52 tokens in the piece rate treatment, which is not significantly different (one-sample t-test: p=0.49) from 400 tokens (earned in the salary treatments).

Finally, the data allow us to explore additional sources of motivation such as task-based incentives, which we undertake in the penultimate section.

III. Variable measurement

The measurement of individuals' effort and their task motivation involves several key decisions. This section describes how we address these measurement issues.

Effort measure

The goal of gaming incentives is to increase effort. Effort varies along several dimensions, however, none of which are easily measured directly. On the intensive margin individuals can exert effort to increase the quantity of output that they produce per period and the quality of the output; on the extensive margin, they can choose to exert effort over longer or shorter periods of time. Our task lends itself to multiple measures of effort: a measure of total effort that gives equal weight to the quality and quantity of output (the sum of words correctly and incorrectly coded); a measure that recognizes only output that meets a minimum quality standard (the number of words correctly coded); and a measure of effort on the extensive margin (the number of rounds individuals choose to spend coding words). Furthermore, the gaming incentives are all meant to increase effort on the intensive margin (the number and accuracy of words coded per round), rather than the extensive margin. Our results are robust to all three measures of effort; however we prefer to report the main results on the extensive margin, but provide additional robustness checks using total effort in the appendix. This is because effort on the extensive margin is most directly under the subjects control, and does not vary by subject ability. In addition, the robustness of results to a measure of effort on the extensive margin boosts the external validity of our findings. Regardless of worker characteristics or the nature of their specific task, additional effort on the extensive margin yields more output. Hence, our findings that gaming incentives increase the willingness of our subjects to undertake the decoding task for more rounds are more likely to translate to workers with other characteristics and engaged in other tasks.

Ability measure

Though ability has only an indirect effect on measured effort on the extensive margin, it directly determines the translation of effort into coded words. All specifications therefore include an exact control for ability. Subjects first familiarized themselves with the decoding task in a practice round. They then undertook the task for an additional two-minute round, during which they received a piece rate of 10 tokens for every word they decoded correctly. The number of correctly coded words in this round constitutes our measure of ability.

Table D.1 in appendix D presents summary statistics and a balance table that tests whether subjects in each treatment are equivalent with respect to observables to subjects in the baseline. Subject gender, age, clarity of instruction, and ability differ at times, mainly due to the small sample. All specifications reported control for these variables, and our results are robust to the inclusion of these controls.

Task motivation measure

The effort effects of gaming incentives depend on individuals' intrinsic motivation to undertake the task. To capture subjects' task motivation, all subjects responded to Ryan's (1982) task evaluation questionnaire during the exit survey. The questionnaire is a part of Intrinsic Motivation Inventory (IMI), which is used in assessing the subjective experiences of participants when developing an activity (Amabile et al. 1994). The interest/enjoyment subscale of IMI serves as a proxy for task motivation. Appendix C lists the questions for the task evaluation questionnaire.

Ideally, we would measure task motivation at the beginning of the experiment so that, by construction, it would have been independent of our treatments. For example, we could have asked the subjects to decode for a round, with no motivational elements present, and then asked how much they enjoyed the decoding task. We did not ask for this at the beginning, but at the exit survey. This may yield the task motivation endogenous to the treatment if subjects responded to the task motivation measure differently by treatment. We do not observe any treatment differences on this measure. The balance table presents tests between the 7 treatments and the baseline, and the coefficients on the treatments are not significantly different from each other. This is because when measuring task motivation, we were careful to word the task enjoyment questions such that they referred repeatedly to the decoding task itself, not to subjects' overall impressions of the activity.

Furthermore, to alleviate these concerns we also develop an alternate measure of task motivation by regressing the treatment dummies on our measure of task motivation and using the residuals from this regression as our alternate measure of motivation. Hence, these residuals are independent of any variation arising from treatment differences. We test whether the results are robust to this alternate measure of motivation and find that our reported results are identical. This robustness check can be found in table D.6 in appendix D. We can therefore also reject the possibility that the positive effect of task motivation on effort arose because the treatment reduced measures of task motivation among those who exerted less effort and increased measures of task motivation among those who exerted more effort.

Survey measures – control variables

The exit survey includes other socio-demographic variables (age, gender, income relative to others, risk preferences, and competitiveness). Risk aversion was measured using subject response to the question "Are you someone who takes risks or do you generally avoid taking risks?" Subjects also responded to the 14 item Competition Index (Smither and Houston, 1992; Houston et al., 2002) designed to measure "a desire to win in interpersonal situations." This item has high reliability and stable properties (Harris and Houston 2010). Generally, our control variables are balanced across treatments (see table D.1), with the exception of gender, age, ability, and clarity of instructions. Wherever possible, we control for each of these variables.

Implementation

The experiment consists of four main parts, depicted in Figure 2. Subjects have a chance to familiarize themselves with the decoding task in a practice round. We then include an "ability measure" prior to the main task, which is incentivized with a piece rate of 10 tokens for every word that subjects decoded correctly in two minutes. Afterwards, the actual task follows. After each round, subjects decide whether to continue the task or not. The experiment concludes with an exit survey.

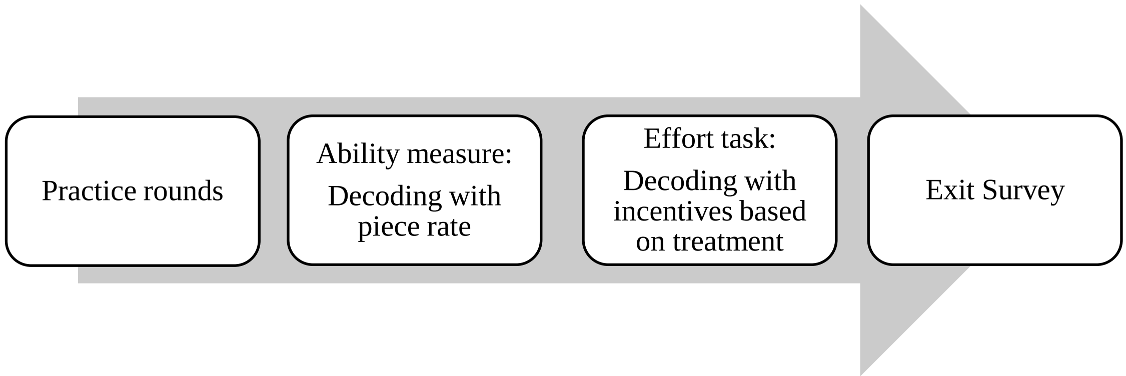


Figure 2: Experiment structure

A total of 362 subjects (undergraduate students at the University of East Anglia) participated in the experiment. Subjects were recruited using the online database system Hroot (Bock et al., 2012). Each subject only participated in a single session of the study. The experiment was programmed and conducted with z-Tree (Fischbacher, 2007). All sessions were run under a single-blind protocol. On average, a session lasted 45 minutes, including the payment, and subjects earned £11 (including the show-up fee).

Before entering the laboratory, subjects drew a ping-pong ball with a number from an opaque bag. They then took their seat in the cubicle with the same number. After sitting, subjects were welcomed and instructed to ask all questions in private, after raising their hand. All instructions were on the computer screen and subjects could proceed at their own pace. Subjects were randomized to treatment within the session, so that subjects were participating in all treatments at the same time. Subjects had to respond to a simple quiz to demonstrate understanding prior to proceeding to the decision-making part of the experiment. After completing the main task, subjects filled out the exit survey. Once they finished, they were instructed to take all their belongings, collect their payment in private at the back of the laboratory, and leave quietly.

Subjects were informed (via the recruitment platform) that the experiment would take an hour of their time. In fact, few sessions lasted that long and subjects finished the experiment in 40 minutes on average. The fact that subjects were free to leave upon completing the experiment might have induced pressure on others to leave earlier than they otherwise might have. However, after choosing to stop the task and before leaving, subjects took an extensive survey to complete the session, which required about ten minutes on average. Hence, it was rare that any one subject left the laboratory before most of the subjects had begun the exit survey.

In addition to the treatments reported in this paper, we conducted one additional treatment with no feedback whatsoever (i.e. subjects received no information on the number of words decoded), as even our baseline treatment has some measure of feedback (subjects are told the number of words they decoded). We conducted this to understand the effects of no feedback (relative to basic feedback) on effort choices. Previous research finds that basic performance feedback has heterogeneous effects and, for some, those effects can be zero or negative. We find that a baseline with no feedback and a baseline with feedback yield statistically indistinguishable levels of effort. We exclude the no feedback condition in our analysis, as that addresses a different question.

To summarize, subjects were asked to engage in the decoding task for as long as they liked. As they decoded the words, the program listed the words at the bottom of the screen.

After each round of decoding, subjects are provided performance feedback, details of which depended on the treatment. At the end of each round, subjects were asked whether they would like to continue with the decoding, or to end the task. The next section presents the main results of the experiment.

IV. Main Results

We first examine the impact of our treatments on the quantity of effort provided. Our primary measure of effort is the number of rounds the subject chose to engage in the task. This is a useful measure because it focuses on effort on the extensive margin, and the choice to continue in the task itself is independent of our subject's performance. Alternatively, one could use alternative measures of effort (number of words coded correctly, or the number of words attempted). The results we report are similar to these alternative measures, though these measures are noisier due to additional variation based on subjects' skill in the task itself. Results using these alternative measures are provided in the appendix (tables D.2, D.3, D.4, and D.5 in appendix D).

Figure 3 provides our main treatment effects. Note that the baseline (an unconditional wage) provides no extrinsic incentives for subjects to engage in the task. Nevertheless, subjects continued this task for 0.88 rounds on average (significantly greater than 0, two tailed t-test p-value = 0.00), with only 40% of the sample stopping after the first (mandatory) round. We also find that the likelihood of continuing the task is significantly correlated with our measure of task motivation (correlation coefficient=0.39, p=0.01), providing some validity to the measure. The first four treatments add each gaming element individually to the baseline task, while the fifth treatment adds all four gaming elements simultaneously. The last two treatments implement a piece rate condition, providing a strong extrinsic incentive to continue the task. The penultimate treatment implements a piece rate only condition, with no gaming elements, while the final treatment adds all gaming elements.

The first gaming element (Points) provides subjects with 50 points per word decoded correctly, adjusted by a multiplier that starts at 1 and rises by 0.1 for every word decoded correctly in a row (to distinguish the points from a simple counter that is used in the baseline). In this treatment, subjects repeated the task 1.12 rounds on average, which is higher but not significantly different from the baseline (2-sample t-test: p=0.42).

The second gaming element (Badges) provides subjects with pre-specified goals and symbolic rewards for achieving the goals. Effort under this treatment increases to 1.36 rounds on average, which is higher, but not significantly different from the baseline (2-sample t-test: p=0.15). The third gaming element (Leaderboard) displays where subjects rank against their peers. This treatment increases effort provided to 1.38 rounds, similar to symbolic rewards but more tightly estimated, yielding a marginally significant increase over the baseline (2-sample t-test: p=0.054).

Of the gaming element treatments, the largest increase comes from the Story/theme element, where subjects are presented with a short story in each round and the task completes words that fit into the story, providing subjects with a purpose to the task itself. Effort increases to 1.70 rounds on average, nearly double the number of rounds in the baseline, yielding a marginally significant increase (2-sample t-test: p=0.052). Thus, of the four gaming elements we test in this paper, competition (through a leaderboard) and purpose (through the use of narratives) yielded higher effort.

One critique of the Story/theme treatment is if the paragraph provided allowed the task to become easier for the subjects, and this added simplicity led to greater effort. However, we note that the words that needed to be de-coded were identical in all treatments and presented in the same order. Furthermore, given the speed required to engage in this task, it is highly unlikely that subjects were able to infer the words from the text itself. The time constraint was substantial: In no treatment, including Story/theme, did subjects come close to decoding all the words (total of 20 per round). The average number of words decoded by subjects across all rounds was 9, with a maximum of 17 ever attained. The maximum achieved in the Story gaming element treatment was 13. Moreover, the number of words decoded in the Story/theme treatment is lower, but statistically indistinguishable from the baseline (p=0.49). Hence, we can reject the supposition that increase in effort provided for the Story/theme treatment is due to the task being simpler.

Next, we turn to the effect of adding all gaming elements on effort. This treatment implements all gaming elements simultaneously, and significantly increases effort. Subjects increase effort to 1.90 rounds on average, yielding a significant increase (2-sample t-test: p=0.001). Hence, the gaming elements together increase effort, but the effects of each element are not additive: the increase from all four gaming elements operating together is marginally higher than the Points gaming element by itself (2-sample t-test: p=0.07), and not significantly different from the other three gaming elements.

Perhaps unsurprisingly, the piece rate treatment, which adds a high-powered incentive to exert effort, significantly increases effort by a large extent. Subjects increase effort provision nearly 10 times, to 10.45 rounds on average (2-sample t-test: p=0.000), and 60% of the sample (33 out of 55 subjects) continuing the task till the maximum possible 13 rounds. However, the main point of interest is the final treatment, which adds all gaming elements to the piece rate pay scheme. Here, effort is certainly higher than the baseline (2-sample t-test: p=0.000) and stands at 10.38 rounds on average. Note, however, that this increase in effort is not significantly different from the piece rate treatment (2-sample t-test: p=0.92). Furthermore, the proportion of subjects hitting the maximum number of rounds is 50% (20 out of 40 subjects), which is not significantly different from the piece rate treatment (2-sample t-test or proportions: p=0.33), nor is the distribution of effort different between the two treatments (Kolgorov-Smirnov test: p=0.92). While there are clearly ceiling effects on effort, the evidence is in line with diminishing returns to multiple sources of motivation, with gaming elements having a significant increase in effort under low extrinsic incentives but having limited effects under high powered extrinsic incentives. That being said, the difference in difference between the gaming treatments under an unconditional wage, and gaming treatments under a piece rate, is not significant (t-test: p=0.19) meaning that we cannot state that the effects of gaming features different under different extrinsic incentive schemes.

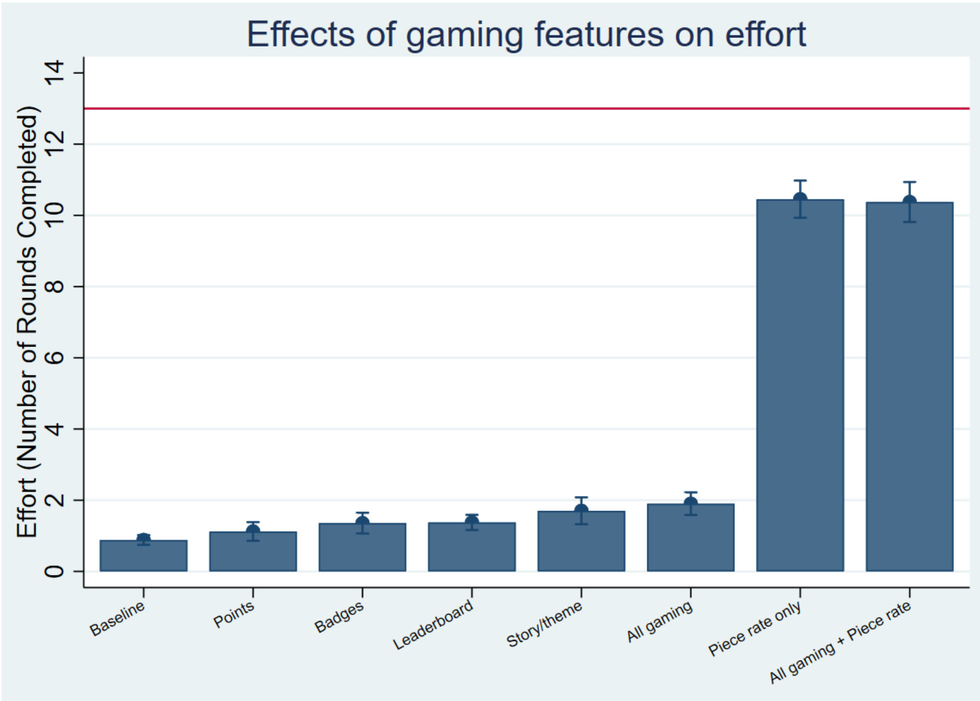


Figure 3: Treatment effects on effort

Table 4 introduces a more formal test of the effects observed in the figure above. We use a Tobit specification to account for censoring in our data (subjects hitting a maximum of 13 rounds beyond the first round), with robust standard errors. Given that we have 8 treatments in total, with 8 hypotheses (7 treatment effects comparing the treatments against the baseline, plus testing the effect of gaming elements in the piece rate treatments), we account for multiple hypothesis testing by reporting Anderson's sharpened q-values (Anderson, 2008, Banerjee et al., 2015, Bryan et al., 2021). This methodology varies the step-down False Discovery Rate (Benjamini et al., 2006) threshold and for each hypothesis calculates the minimum q-value for which it would be rejected. These are reported for the treatment effects in table 4 (in square brackets). Correcting from multiple hypothesis tests does not affect our results. Finally, the results we report are robust to using alternative measures of effort. These are reported in appendix D (tables D.2 and D.3 for words coded correctly, and words attempted respectively).

Model I of Table 4 presents the pure treatment effects of the experiment, and corresponds directly with figure 3. The Tobit specification accounts for censoring in the data, but otherwise, the results are the same as in the figure: leaderboards and story/theme gaming elements have a marginally significant positive impact on effort, while gaming elements together have a positive and significant impact. Furthermore, the piece rate treatments have a large positive impact, but gaming elements do not increase effort under piece rates.

Model II adds controls for gender, age, and subject ability (using data from an earlier piece rate round). Model III adds controls for task motivation and preferences for competition. Note that task motivation has a significant positive impact on effort provided ($p<0.05$ across all specifications), indicating that effort is higher for those that report higher levels of intrinsic motivation, giving further credence to the motivation measure. Finally, model IV adds controls for risk preferences, relative income, and clarity of instructions. Our main treatment effects remain largely unchanged across all these specifications.

Table 4: Treatment effects on effort provision (Rounds continued)

Dependent Variable: Quantity of effort provided (Rounds continued)				
	I	II	III	IV
Gaming feature 1 (Points)	0.238	0.152	0.092	0.089
	(0.29)	(0.32)	(0.32)	(0.32)
	[0.217]	*[0.370]*	*[0.407]*	*[0.439]*
Gaming feature 2 (Badges)	0.472	0.402	0.290	0.320
	(0.32)	(0.33)	(0.33)	(0.33)
	[0.091]	*[0.177]*	*[0.231]*	*[0.197]*
Gaming feature 3 (Leaderboard)	0.491*	0.436	0.421*	0.458*
	(0.25)	(0.27)	(0.25)	(0.27)
	[0.055]	*[0.113]*	*[0.104]*	*[0.093]*
Gaming feature 4 (Story/Theme)	0.846**	0.797*	0.737*	0.757*
	(0.42)	(0.42)	(0.42)	(0.43)
	[0.055]	*[0.074]*	*[0.104]*	*[0.093]*

All Gaming features	1.031***	0.972***	0.865***	0.893***
	(0.35)	(0.34)	(0.33)	(0.33)
	[0.009]	[0.011]	[0.019]	[0.017]
Piece rate	10.67***	10.61***	10.54***	10.45***
	(0.73)	(0.74)	(0.74)	(0.74)
	[0.001]	[0.001]	[0.001]	[0.001]
Piece rate + All gaming features	10.37***	10.29***	10.18***	10.20***
	(0.76)	(0.76)	(0.73)	(0.74)
	[0.001]	[0.001]	[0.001]	[0.001]
Gaming features under piece rate (q-values)	[0.241]	[0.398]	[0.407]	[0.439]
Gender		0.182	0.311	0.334
1 = Female		(0.31)	(0.33)	(0.33)
Age (in years)		-0.024	-0.023	-0.025
		(0.05)	(0.05)	(0.05)
Ability in task		0.022	-0.006	-0.034
		(0.05)	(0.05)	(0.06)
Task motivation			0.452**	0.423**
5 = Motivated			(0.18)	(0.18)
Competitiveness			0.410	0.351
5 = Highly competitive			(0.27)	(0.28)
Risk preferences				0.003
5 = Risk-seeking				(0.17)
Income (relative to others)				-0.093
5 = Far above average				(0.16)
Clarity of instructions				0.308*
5 = Always clear				(0.18)
Constant	0.884***	1.119	-1.398	-1.913
	(0.14)	(1.23)	(1.42)	(1.46)
Observations	362	362	362	362
Log likelihood	-810.8	-810.5	-806.4	-805.0
P	0.000	0.000	0.000	0.000
Pseudo R-squared	0.210	0.211	0.215	0.216
Right censors	56	56	56	56

Note: Tobit specifications with robust standard errors in parentheses. Anderson's sharpened q-values in [] to account for multiple hypothesis testing. * 10%, ** 5%, *** 1% significance level.

Broadly the results are consistent with those observed in figure 3. Of the gaming elements, leaderboards and story/theme exhibit small increases in effort. Gaming elements operating together yield significant increases in effort. Our piece rate treatments yield a large increase in effort overall, while the addition of gaming elements does not significantly increase effort, in line with diminishing returns to multiple sources of motivation.[12] These results are robust to alternative measures of effort. Finally, we note that our measure of task motivation is correlated with effort across all specifications.

V. Exploratory Results

An interesting aspect of the results is the strong correlation between effort and intrinsic (task) motivation. The relationship between these two variables is significant both in the baseline condition, and across all treatments. In this section we present results exploring the treatment effects and how they differ across motivated and unmotivated groups.

Our measure of motivation is elicited by a survey at the end of the experiment. We ask subjects to respond to the following questions:

- While I was working on the decoding task I was thinking about how much I enjoyed it.
- I found the decoding task very interesting.
- Doing the decoding task was fun.
- I enjoyed doing the decoding task very much.
- I thought the task was very boring. (reversed)
- I thought the decoding task was very interesting.
- I would describe the task as very enjoyable.

The questions all refer to the decoding task, which was neutrally presented to the subjects at the start of the experiment. Our measure of motivation is a simple average of the response to the survey items above (Cronbach's $\alpha = 0.9229$). That is, all subjects were given instructions on the decoding task and asked to practice the task prior to any treatment manipulations. The questions all refer to the decoding task itself, without specifying the treatment manipulations. Hence, while the subjects are asked to respond about the decoding

[12] Since many of the treatment dummies yield insignificant effects, it is useful to know the effect sizes for each treatment dummies based on an OLS specification with the same controls as in Table 4, Model IV. We find that 48.4% of the variation in the number of rounds continued is explained by the piece rate treatment, and 43.7% of the variation is explained by the piece rate with gaming features treatment. By contrast, our gaming features treatments in the absence of high extrinsic incentives explain relatively less: All gaming features explain 0.8% of the variation in the data, while the effect size is the largest for Stories (0.55%) followed by Leaderboards (0.2%), Badges (0.11%) and finally, Points (0.01%). Furthermore, when comparing the two high extrinsic incentive treatments with each other, we find that adding all gaming features under high extrinsic incentives explains just 0.01% of the variation in the data, yielding a much smaller effect size compared to the low extrinsic incentive treatments (0.8%).

task itself, it is not clear whether their responses are also affected by the treatment, yielding concerns about endogeneity. To partially alleviate these concerns, we note that task motivation does not differ across treatments in our balance table (D.1) with the exception of the piece rate treatment with all gaming elements, where motivation is marginally significant and positive relative to the baseline. Furthermore, to account for this potential source of endogeneity, we run a simple OLS regression of treatment dummies on effort (essentially the same model as presented in the balance table) and capture the residuals. This is our alternative measure of motivation that is independent of variation arising from treatment, which we use to check for robustness. The results using the alternative measure are in table D.6 in appendix D. Note that the results we present below are identical to those using this alternative measure of motivation.

Figure 4 displays the average effort levels (number of rounds continued) for each treatment, broken out by task motivation. The figure splits the sample at the mean of task motivation for the entire sample, hence the less motivated group contains subjects whose task motivation was at or below the average motivation for all subjects participating in the experiment. Effort in the baseline condition (the first two bars of Figure 4) is marginally greater among more task-motivated subjects (two-sample t-test: $p=0.08$). The same pattern (more motivated subjects exerting higher effort) is true for all treatments except for two: the treatment with Story/theme gaming element, and the pure piece rate treatment. In both these cases, subjects that are more task motivated exert less effort, though the difference is not significant. When compared to the baseline, the *less* task-motivated subjects in the Story/theme gaming treatment exert significantly more effort than the less task -motivated in the baseline (p=0.03). More task-motivated subjects are not significantly different across these two treatments (p=0.60), indicating that Story/theme gaming treatment increases effort among those who are less motivated by the task.

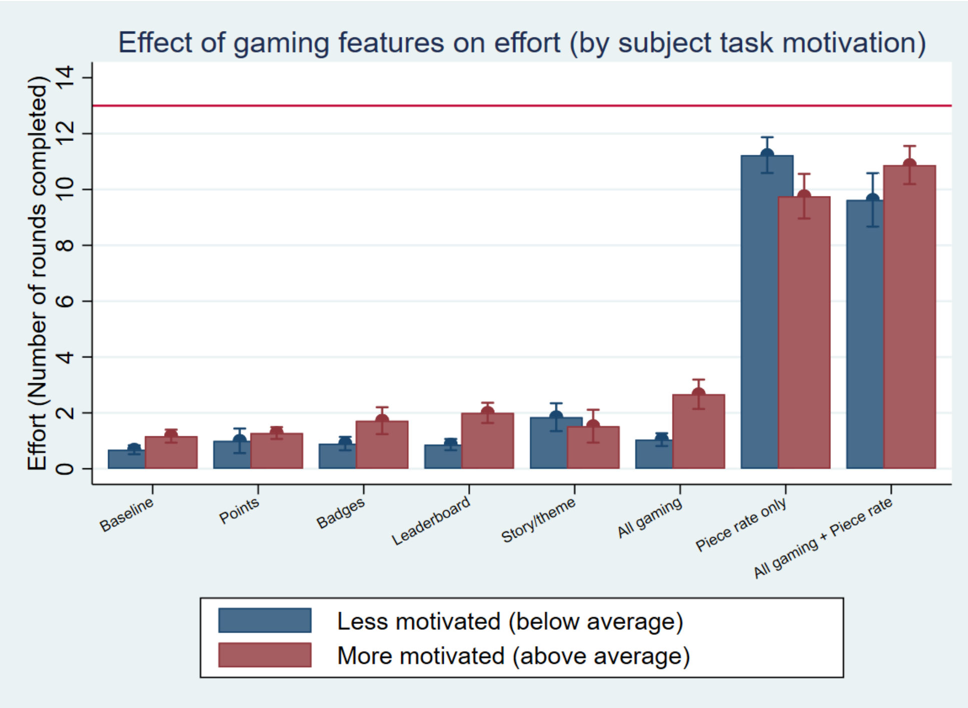


Figure 4: Treatment effects on effort (heterogeneity by task motivation)

Table 5 presents a more formal test of what we observe in the figures. Using the same dependent variable for effort (rounds continued), we interact treatment dummies with our variable for task motivation. Model 1 presents just the treatment dummies, task motivation, and the interactions between the treatment and motivation. Model II adds controls for gender, age, ability, competitiveness, risk preferences, relative income, and clarity of instructions. The table reports coefficients from Tobit models to account for censoring. What is immediately clear from this is that the interaction term between the Story/theme treatment and task motivation is negative and significant (p=0.03). Moreover, this term is significantly different from the uninteracted task motivation variable (p=0.01) indicating that the slope of the relationship between task motivation and effort is flatter in the Story/theme treatment relative to the baseline.

Table 5: Treatment and task motivation interactions

Dependent Variable: Quantity of effort provided (Rounds continued)		
	I	II
Gaming feature 1 (Points)	0.232	0.228
	(0.71)	(0.82)

Gaming feature 2 (Badges)	0.383	0.463
	(0.63)	(0.69)
Gaming feature 3 (Leaderboard)	-0.626	-0.398
	(0.62)	(0.77)
Gaming feature 4 (Story/Theme)	2.270**	2.915***
	(0.91)	(1.06)
All Gaming features	-1.166	-0.972
	(0.79)	(0.80)
Piece rate	12.88***	12.95***
	(2.49)	(2.45)
Piece rate + All gaming features	5.834**	5.919**
	(2.57)	(2.62)
Task motivation	0.367***	0.387***
5 = Motivated	(0.13)	(0.15)
Task motivation X	-0.010	-0.030
Gaming feature 1 (Points)	(0.23)	(0.27)
Task motivation X	-0.006	-0.032
Gaming feature 2 (Badges)	(0.24)	(0.26)
Task motivation X	0.376	0.305
Gaming feature 3 (Leaderboard)	(0.24)	(0.28)
Task motivation X	-0.489*	-0.719**
Gaming feature 4 (Story/Theme)	(0.28)	(0.33)
Task motivation X	0.664**	0.591*
All Gaming features	(0.30)	(0.31)
Task motivation X	-0.747	-0.804
Piece rate	(0.80)	(0.79)
Task motivation X	1.397*	1.358
Piece rate + All gaming features	(0.82)	(0.84)
Gender		0.286
1 = Female		(0.33)
Age (in years)		-0.024
		(0.05)
Ability in task		-0.050
		(0.06)
Competitiveness		0.261
5 = Highly competitive		(0.27)
Risk preferences		0.067
5 = Risk-seeking		(0.17)
Income (relative to others)		-0.153
5 = Far above average		(0.15)
Clarity of instructions		0.321*
5 = Always clear		(0.18)
Constant	-0.179	-1.443
	(0.33)	(1.40)
Observations	362	362
Log likelihood	-802.2	-799.2
P	0.000	0.000

Pseudo R-squared	0.219	0.222
Right censors	56	56

Note: Tobit specifications, robust standard errors in parentheses. * 10%, ** 5%, *** 1% significance level.

The analysis indicates that the Story/theme gaming element has a large, positive effect on effort, but especially on *less* task-motivated individuals. Among those who were least interested in decoding words, the presence of the Story gaming element was highly motivating. However, for those motivated by the task itself, the Story gaming element had essentially no effect and their effort was not significantly different than in the baseline condition. This suggests that the positive effect of the gaming element treatment is operating through the least motivated subjects. As before, these results are robust to alternative measures of effort (see tables D.4 and D.5 in appendix D), as well as a revised measure of effort using the residuals of a simple model with treatment dummies (table D.6 in appendix D), to account for concerns about endogeneity in the motivation measure.

VI. Conclusions

In recent years, many organizations have experimented with introducing gaming elements to motivate workers/users. However, organizations have little empirical guidance on whether these gaming incentives work to increase effort/engagement, how these gaming incentives interact, both with each other, and with more traditional incentives, and whether the effects vary by intrinsic incentives. Our experiment provides the first systematic evidence on this topic, using a standard real effort task in a controlled laboratory environment. We report three main results that can help organizations understand how gaming incentives can motivate effort given their current incentive structures. First, gaming incentives can increase effort, although the size of the effect remains small, and with some types of incentives more motivating than others (leaderboards and stories). Second, gaming incentives exhibit diminishing returns: they increase effort under low extrinsic incentives, but effort is not different under high extrinsic incentives (and may even be lower). Finally, we find that only one type of gaming incentive (the story/theme) increases effort in the less motivated subsample, indicating that different types of gaming incentives trigger effort in different kinds of workers, and some even increasing effort among those that do not enjoy the task itself.

The fact that gaming incentives increase effort is not too surprising given the high levels of usage in contemporary organizations. However, there is a dearth of empirical evidence, with gaming incentives sometimes yielding positive effects, and sometimes not (Hamari, Koivisto, and Sarsa, 2014). We offer the first evidence why this might be the case by analyzing the interaction between gaming incentives and extrinsic/intrinsic motives to exert effort. Our results are important for organizations seeking to implement gaming incentives by shining a light on the importance of existing incentive schemes, which can help to understand the heterogeneity in outcomes as a result of implementation. Finally, we note that these results are economically meaningful. They can generate considerable increases in effort and engagement at relatively lower marginal costs. Hence, the implication for the efficiency and operation of firms can be quite large.

References

Amabile, T.M., Hill, K.G., Hennessey, B.A. and Tighe, E.M., 1994. The Work Preference Inventory: assessing intrinsic and extrinsic motivational orientations. *Journal of personality and social psychology,* 66(5), p.950.

Ariely, D., Bracha, A. and Meier, S. (2009). "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially." *The American Economic Review* 99(1): 544-555.

Ariely, D., Kamenica, E. and Prelec, D. (2008). "Man's search for meaning: The case of Legos." *Journal of Economic Behavior & Organization* 67(3): 671-677.

Ashraf, N., Bandiera, O. and Jack, B.K. (2014). "No margin, no mission? A field experiment on incentives for public service delivery." *Journal of Public Economics* 120: 1-17.

Azmat, G. and Iriberri, N. (2010). "The importance of relative performance feedback information: Evidence from a natural experiment using high school students." *Journal of Public Economics* 94(7): 435-452.

Bandiera, O., Barankay, I. and Rasul, I. (2007). "Incentives for managers and inequality among workers: Evidence from a firm-level experiment." *The Quarterly Journal of Economics* 122(2): 729-773.

Banuri, S., Keefer, P., and Damien de Walque (2021). "Love the job... or the patient? Task vs. mission-based motivations in healthcare." Mimeo.

Barankay, I. (2012) "Rank incentives: Evidence from a randomized workplace experiment." Unpublished manuscript.

Bellemare, Charles and Bruce Shearer (2009). "Gift giving and worker productivity: Evidence from a firm-level experiment." *Games and Economic Behavior* 67: 23-244.

Blanes i Vidal, J. and Nossol, M. (2011). "Tournaments without prizes: Evidence from personnel records." *Management Science* 57(10): 1721-1736.

Bloom, N. and Van Reenen, J. (2007). "Measuring and explaining management practices across firms and countries." *The Quarterly Journal of Economics* 122(4): 1351-1408.

Bock, O., Nicklisch, A. and Baetge, I., (2012). hroot: Hamburg registration and organization online tool. H-Lab Working Paper (1).

Bradler, C., Dur, R., Neckermann, S., & Non, A. (2016). "Employee recognition and performance: A field experiment." *Management Science* 62(11): 3085-3099.

Carpenter, J. and Myers, C.K. (2010). "Why volunteer? Evidence on the role of altruism, image, and incentives." *Journal of Public Economics* 94(11): 911-920.

Casas-Arce, P. & Martínez-Jerez, F. A. (2009). "Relative performance compensation, contests, and dynamic incentives." *Management Science* 55: 1306-1320.

Cassar, Lea and Stephan Meier (2018). "Nonmonetary Incentives and the Implications of Work as a Source of Meaning." *Journal of Economic Perspectives* 32(3): 215-238.

Corgnet, Brice, Joaquín Gómez-Miñambres, and Roberto Hernán-Gonzalez. "Goal setting and monetary incentives: When large stakes are not enough." *Management Science* 61, no. 12 (2015): 2926-2944.

Dato, S. and Nieken, P., 2014. Gender differences in competition and sabotage. Journal of Economic Behavior & Organization, 100, pp.64-80.

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

Deci, E.L. and Ryan, R.M., (1985). "The general causality orientations scale: Self-determination in personality." *Journal of research in personality* 19(2): 109-134.

DellaVigna, Stefano (2017). "Structural Behavioral Economics." Mimeo, University of California, Berkeley.

Deterding, Sebastian. "Situated motivational affordances of game elements: A conceptual model." In Gamification: Using game design elements in non-gaming contexts, a workshop at CHI, vol. 10, no. 1979742.1979575. 2011.

Deterding, Sebastian, Dan Dixon, Rilla Khaled, and Lennart Nacke. "From game design elements to gamefulness: defining" gamification"." In Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments, pp. 9-15. 2011.

Dubey, P. K. & Geanakoplos, J. (2005). "Grading in games of status: Marking exams and setting wages." Cowles Foundation Discussion Paper No. 1544, Yale University.

Dubey, P. and Geanakoplos, J. (2010). "Grading exams: 100, 99, 98,… or A, B, C?." *Games and Economic Behavior* 69(1): 72-94.

Ederer, F. (2010). "Feedback and motivation in dynamic tournaments." *Journal of Economics & Management Strategy* 19(3): 733-769.

Eriksson, T., Poulsen, A. and Villeval, M.C. (2009). "Feedback and incentives: Experimental evidence." *Labour Economics* 16(6): 679-688.

Erkal, N., Gangadharan, L. and Nikiforakis, N. (2011). "Relative earnings and giving in a real-effort experiment." *The American Economic Review* 101(7): 3330-3348.

Falk, A., Gächter, S. and Kovács, J. (1999). "Intrinsic motivation and extrinsic incentives in a repeated game with incomplete contracts." *Journal of Economic Psychology* 20(3), pp.251-284.

Fehr, E. and Falk, A. (1999). "Wage rigidity in a competitive incomplete contract market." *Journal of Political Economy* 107(1): 106-134.

Fischbacher, U. (2007). "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics* 10(2): 171-178.

Frey, Bruno S., and Felix Oberholzer-Gee (1997). "The cost of price incentives: An empirical analysis of motivation crowding-out." *The American Economic Review* 87(4): 746-755.

Friedrichsen, J., and Engelmann, D. (2017). "Who Cares about Social Image?" (No. 1634). DIW Berlin, German Institute for Economic Research.

Gibson, J. J. (1977). The theory of affordances. *Hilldale, USA*, 1(2), 67-82.

Hamari, J. (2013). "Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service." *Electronic Commerce Research and Applications* 12(4): 236-245.

Hamari, Juho, Jonna Koivisto, and Harri Sarsa. "Does gamification work?--a literature review of empirical studies on gamification." In 2014 47th Hawaii international conference on system sciences, pp. 3025-3034. Ieee, 2014.

Harris, P. B., & Houston, J. M. (2010). "A reliability analysis of the revised competitiveness index.: *Psychological Reports* 106(3): 870-874.

Herzberg, F. I. (1966). *Work and the nature of man*. Oxford: World Publications

Houston, J. M., Harris, P. B., McIntire, S., & Francis, D. (2002) "Revising the Competitiveness Index." *Psychological Reports* 90: 31-34.

Hsiaw, A. (2013). Goal-setting and self-control. *Journal of Economic Theory*, 148(2), 601-626.

Johnson, D.B. and Ramalingam, A., 2016. Wage compression and manager inequality aversion (No. 16-13). School of Economics, University of East Anglia, Norwich, UK..

Klock, Ana Carolina Tomé, Isabela Gasparini, Marcelo Soares Pimenta, and Juho Hamari. "Tailored gamification: A review of literature." International Journal of Human-Computer Studies 144 (2020): 102495.

Kluger, A.N. and DeNisi, A. (1996). "The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory." *Pscyhological Bulletin* 119(2):254-284.

Koch, A. K., & Nafziger, J. (2011). Self-regulation through goal setting. *Scandinavian Journal of Economics*, 113(1), 212-227.

Kosfeld, Michael and Susanne Neckermann (2011). "Getting More Work for Nothing? Symbolic Awards and Worker Performance." *American Economic Journal: Microeconomics* 3(August): 86-99.

Krath, Jeanine, Linda Schürmann, and Harald FO Von Korflesch. "Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning." Computers in Human Behavior 125 (2021): 106963.

Laughlin, P.R., Bonner, B.L. and Miner, A.G., 2002. Groups perform better than the best individuals on letters-to-numbers problems. Organizational Behavior and Human Decision Processes, 88(2), pp.605-620.

Laughlin, P.R., Hatch, E.C., Silver, J.S. and Boh, L., 2006. Groups perform better than the best individuals on letters-to-numbers problems: effects of group size. Journal of Personality and social Psychology, 90(4), p.644.

List, J.A. and Rasul, I. (2011). "Field experiments in labor economics." *Handbook of Labor Economics* 4:103-228.

Lizzeri, A., Meyer, M.A. and Persico, N., (2002). "The incentive effects of interim performance evaluations." University of Pennsylvania, Center for Analytic Research in Economics and the Social Sciences.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*. Prentice-Hall, Inc.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist*, 57(9), 705.

McDonald, I.M., Nikiforakis, N., Olekalns, N. and Sibly, H., 2013. Social comparisons and reference group formation: Some experimental evidence. Games and Economic Behavior, 79, pp.75-89.

McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. New York: Penguin.

Mekler, Elisa D., Florian Brühlmann, Klaus Opwis, and Alexandre N. Tuch. "Do points, levels and leaderboards harm intrinsic motivation? An empirical analysis of common gamification elements." In Proceedings of the First International Conference on gameful design, research, and applications, pp. 66-73. 2013.

Neitzel, J. and Sääksvuori, L., (2013). "Normative Conflict and Cooperation in Sequential Social

Dilemmas." Verein für Socialpolitik/German Economic Association.

Nikiforakis, N., Noussair, C.N. and Wilkening, T., 2012. Normative conflict and feuds: The limits of self-enforcement. Journal of Public Economics, 96(9-10), pp.797-807.

Norman, D. A. (1999). Affordance, conventions, and design. *interactions*, 6(3), 38-43.

Porter, L. W. and E. E. Lawler (1968). *Managerial attitudes and performance.* Homewood, IL: Irwin-Dorsey.

Ryan, R. M. (1982). "Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory." *Journal of personality and social psychology* 43(3): 450-461.

Ryan, R. M. and E. L. Deci (2000). "Intrinsic and extrinsic motivations: Classic definitions and new directions." *Contemporary educational psychology* 25(1): 54-67.

Sailer, Michael, and Lisa Homner. "The gamification of learning: A meta-analysis." Educational Psychology Review 32, no. 1 (2020): 77-112.

Saleem, Awaz Naaman, Narmin Mohammed Noori, and Fezile Ozdamli. "Gamification applications in E-learning: A literature review." Technology, Knowledge and Learning 27, no. 1 (2022): 139-159.

Smither, R. D., & Houston, J. M. (1992) "The nature of competitiveness: construction and validation of the Competitiveness Index." *Educational and Psychological Measurement* 52: 407-418.

Staw, B. M. (1977). "Motivation in organizations: Toward synthesis and redirection." *New directions in organizational behavior* 1: 54-95.

Suvorov, A., & Van de Ven, J. (2008). Goal setting as a self-regulation mechanism. Available at SSRN 1286029.

Zhang, P. (2008). Motivational affordances: Reasons for ICT design and use. *Communications of the ACM*, 51(11), 145-147.

Appendix A: Instructions and screenshots

Table A.1: Instructions for the main task

Instructions page 1
FINAL PHASE Thank you for completing Phase 1. This is the second and final phase, after which you will be asked to complete a short survey, and then will be paid your earnings for the session. In this phase, you will engage in the decoding task again. However, the task is a little different this time. You will be paid 400 tokens for participating in this phase. This amount is unconditional. This means that you will be paid 400 tokens regardless of your decisions in the final phase. Once this phase is complete, you will be asked to complete a short survey, and then will be paid and be free to leave. Please click "continue" to continue the instructions.

Instructions page 2
FINAL PHASE For the final phase, you will engage in the decoding task again. However, there are a number of differences. The major difference between this phase and previous ones is that the words are not random letters, but actual words. The second difference is that you will now engage in the decoding task for 2 minutes (120 seconds) per round. After the first round, you can choose to end the phase. Once you have completed the first round, you will be given feedback on your performance, and then asked whether you would like to stop the phase or to continue for another round. Please click "continue" to continue the instructions.

Instructions page 3
FINAL TASK If you choose to continue, you will be asked to engage in the decoding task for another 2 minutes. You can choose to continue the phase for as many rounds as you like. If you choose to end the task, you will be asked to start the survey, at the end of which you will be paid for the session and then will be free to leave. It is important to note that you will be paid 400 tokens no matter how long you engage in this phase, or how many words you decode correctly. You are free to end the phase at the end of any round after the first. As soon as you end the phase, you will be given the survey to complete and then paid and free to leave. Please click "continue" to continue the instructions.

Instructions page 4
FINAL TASK If you choose to continue, you will be asked to engage in the decoding task for another 2 minutes. You can choose to continue the phase for as many rounds as you like. If you choose to end the task, you will be asked to start the survey, at the end of which you will be paid for the session and then will be free to leave. It is important to note that you will be paid 400 tokens no matter how long you engage in this phase, or how many words you decode correctly. You are free to end the phase at the end of any round after the first. As soon as you end the phase, you will be given the survey to complete and then paid and free to leave. Please click "continue" to continue the instructions.

Instructions – Story/theme treatment
FINAL TASK Each round contains a short story with certain key words missing. You task is to decode the words to complete the story. Each round contains a different short story, which have been taken from different cultures. As you decode the words, the program will list them at the bottom of the screen. At the end of each round, you will get the story with the words you have decoded filled in. Note that the amount of the story that is revealed depends on the number of words that you decode correctly. The more words you decode, the more of the story you get at the end of the round. Please click "continue" to continue the instructions.

Table A.2: Instructions for the effort task

Instructions – Points Systems treatment

FINAL TASK

In this round there are a number of ways you will be able to track your performance.

There is a points system that is used to keep track of your performance in each round. For each word you decode correctly, you will earn 50 points.

In addition to this, there is a multiplier that starts at 1, and increases by 0.1 for each word you decode correctly in a row. If you make a mistake, the multiplier will reset to 1. Therefore, the more accurate you are, the higher your points will be.

Please click "continue" to continue the instructions.

Instructions – Leaderboards treatment

FINAL TASK

Other students at UEA have participated in similar sessions earlier. They have been ranked according to their highest score in this phase.

Based on the number of words you decode correctly in this phase, you will be assigned a rank and place on the leaderboard. Note that only the highest score you achieve will be used to rank you. No identifying information will be used: You will be anonymously added to the leaderboard at the end of this session.

Also note that others in the session will be added to the leaderboard as well, but only after this session. Therefore, the ranking information you receive will not include other players in the session today.

Please click "continue" to continue the instructions.

Instructions – Quiz – All gaming incentives treatment

1. This phase will continue for four rounds.	○ True ○ False
2. I can choose to end the phase at the end of each round.	○ True ○ False
3. There is a maximum of 20 rounds in this phase.	○ True ○ False
4. Ending the phase means that I will be asked to complete the final survey and then will be free to leave.	○ True ○ False
5. I can continue this phase for as many rounds as I wish.	○ True ○ False
6. I will be paid based on the number of words decoded correctly.	○ True ○ False
7. Each round contains a short story with key words that I have to decode.	○ True ○ False
8. For each word I decode correctly, I will earn points.	○ True ○ False
9. The points multiplier depends on how many words I decode correctly in a row.	○ True ○ False
10. My highest score will be anonymously added to the leaderboard.	○ True ○ False
11. If I code 19 words or more, I will earn the titanium medal.	○ True ○ False

Table A.3: Effort screenshots

Baseline

All gaming incentives treatment

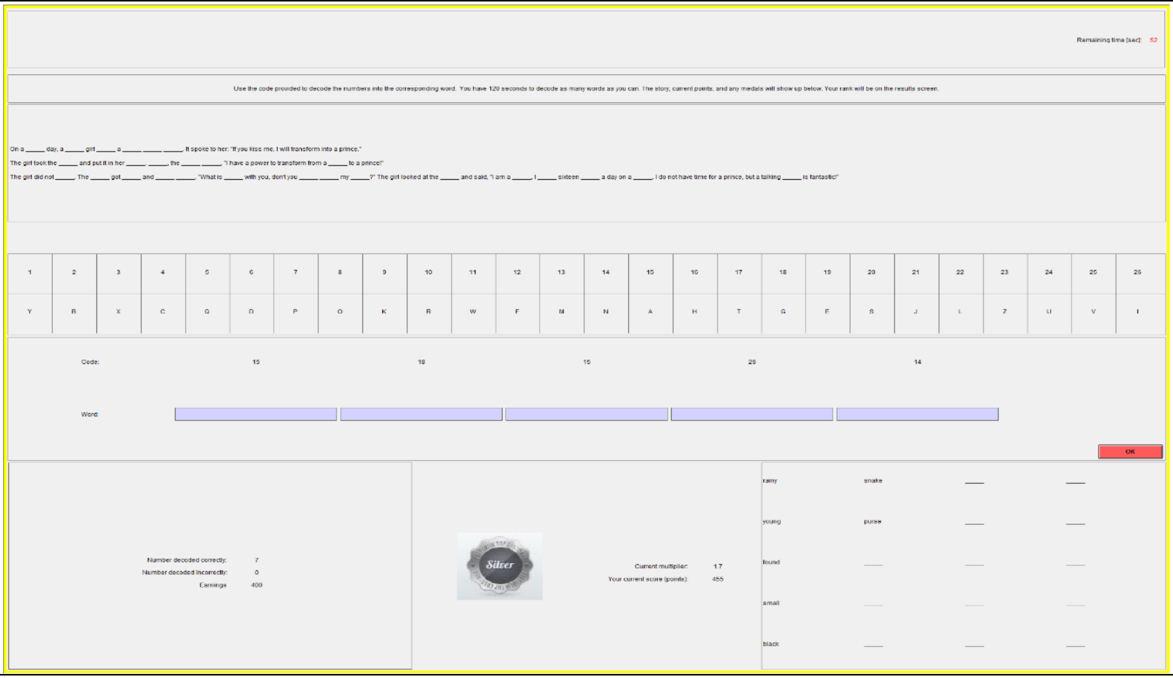
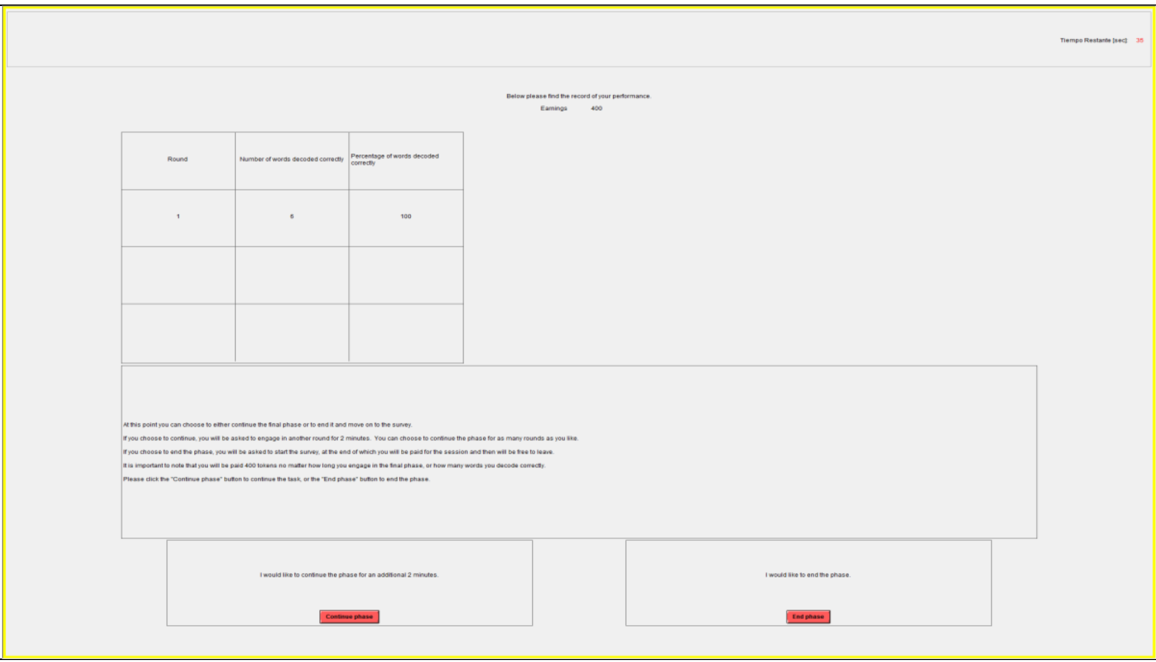
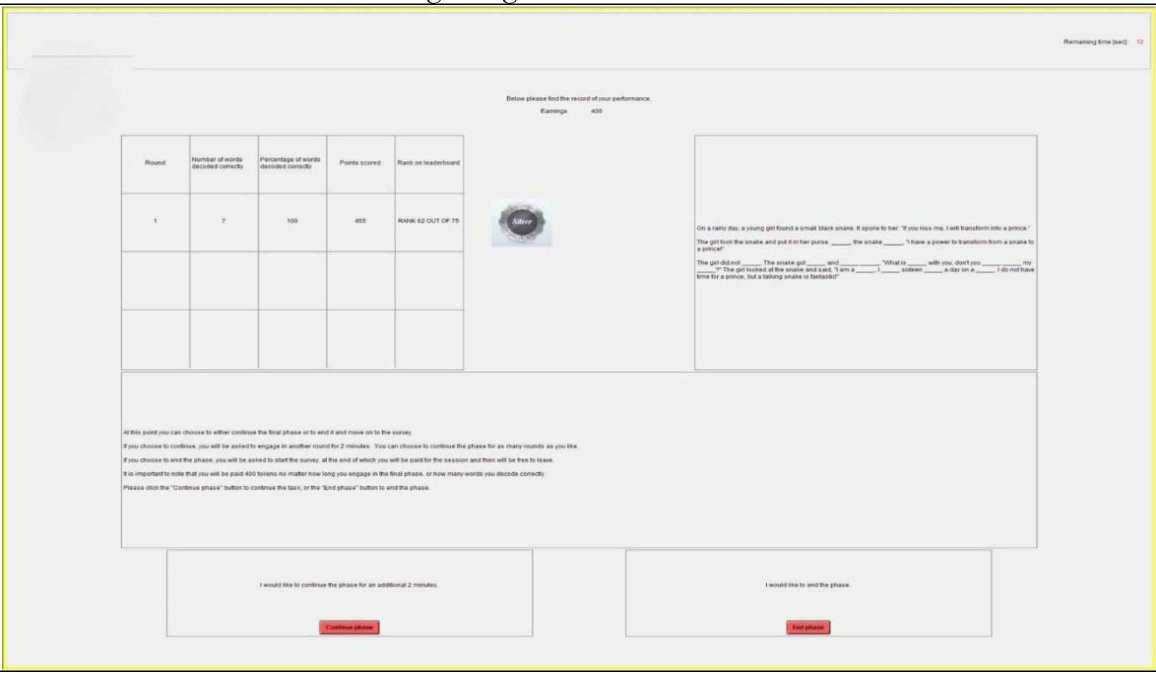


Table A.4: Decision and feedback screenshots

Baseline



All gaming incentives treatment



Appendix B: Intrinsic Motivation Inventory survey questions

Intrinsic Motivation Inventory – Task Evaluation

* Note: highlighted questions were used to measure intrinsic motivations (task evaluation/enjoyment)

1="Strongly disagree"; 2="Disagree"; 3="Neutral"; 4="Agree"; 5="Strongly agree"

1. While I was working on the decoding task I was thinking about how much I enjoyed it.
2. I did not feel at all nervous about doing the decoding task.
3. I think I am pretty good at the decoding task.
4. I found the decoding task very interesting.
5. I felt tense while doing the decoding task.
6. I think I did pretty well compared to other students.
7. Doing the decoding task was fun.
8. I felt relaxed while doing the decoding task.
9. I enjoyed doing the decoding task very much.
10. I am satisfied with my performance at the decoding task.
11. I was anxious while doing the decoding task.
12. I thought the task was very boring. (r)
13. I felt pretty skilled at this task.
14. I thought the decoding task was very interesting.
15. I felt pressured while doing the decoding task.
16. I would describe the task as very enjoyable.
17. After working at this task for a while, I felt pretty competent.

Appendix C: Theoretical model

Basic performance feedback only reminds players of the effort they undertook. Its effects on effort may be enhanced by making goal achievement more interesting or "fun", essentially increasing task motivation; by triggering other intrinsic motivations, such as competitiveness; or by giving greater "purpose" to the task (that is, increasing motivational affordances). If individuals not only accumulate points, but do so in interesting ways (for example, at different rates that depend on previous performance), they might exert more effort. If they receive symbolic rewards, such as gold stars, they may exert more effort than if they merely achieved their goals. Relative performance feedback (for example, through Leaderboards) introduces competitive motivations to increase effort, even when only feedback receivers know their position on the leaderboard. Finally, with the introduction of a story/theme, increasing effort advances a broader narrative that reveals itself more completely when individuals exert more effort. A desire to contribute to this broader mission also creates additional incentives to exert effort.

Underlying all notions of performance feedback is the premise that feedback in its different forms affects utility from undertaking effort. However, the mechanisms of those effects may vary. Motivation from feedback could depend on task motivation – individuals may react less to feedback about their performance when they are uninterested in the task or, as in the case of the story/theme treatment, gaming incentive inspires the most effort among the least task motivated. Moreover, the effects of feedback could be negative for individuals who dislike gaming incentives.

We take an agnostic approach to the contribution of gaming incentives and task motivation to utility. Their contribution to utility is given by $g(\delta_i, \theta_i)$, where $g_{\delta,} g_\theta > 0$ when gaming incentive δ and task θ are motivating (increase utility), otherwise they are less than zero; and where $g_{\delta\theta} > 0$ when task and gaming incentive each increase the motivational effects of the other. In addition, we assume that utility is separable in income unrelated to effort; effort-dependent intrinsic and extrinsic compensation; and the costs of effort. We can then write the utility from effort and pecuniary compensation as

$$u_i = w_f + \left(w_p + g(\delta_i, \theta_i)\right) ln(e_i) - \frac{e_i^\gamma}{\gamma} \qquad (1)$$

Worker utility rises with their unconditional wage, w_f, independently of their effort. To ensure that the intrinsic and extrinsic rewards of effort increase with effort, but at a declining rate, these rewards depend on $ln(e_i)$.[13] The contribution of effort-linked pecuniary incentives to utility, w_p, is separable from the contribution of intrinsic motivations, $g(\delta_i, \theta_i)$. Following DellaVigna (2017) and Bellemare and Shearer (2009), the cost of effort is given by $\left(\frac{e_i^\gamma}{\gamma}\right)$, where γ is the value of effort and the cost of effort increases in effort (i.e. $\gamma > 1$).

Maximizing utility with respect to effort yields an expression that defines optimal effort $\frac{1}{e_i^*}\left(w_p + g(\delta_i, \theta_i)\right) - e_i^{*\gamma-1} = 0 => \left(w_p + g(\delta_i, \theta_i)\right)^{\frac{1}{\gamma}} = e_i^*$. From this we derive key comparative statics. If each type of motivation enters positively in an individual's utility, motivation increases effort:

$$\frac{\partial e_i^*}{\partial w_p} = \frac{1}{\gamma}\left(w_p + g(\delta_i, \theta_i)\right)^{\frac{1}{\gamma}-1} > 0;$$

$$\frac{\partial e_i^*}{\partial \theta_i} = \frac{1}{\gamma}\left(w_p + g(\delta_i, \theta_i)\right)^{\frac{1}{\gamma}-1} g_\theta > 0;$$

$$\frac{\partial e_i^*}{\partial \delta_i} = \frac{1}{\gamma}\left(w_p + g(\delta_i, \theta_i)\right)^{\frac{1}{\gamma}-1} g_\delta > 0.$$

---

13 An alternative formulation is $u_i = w_f + ln\left[\left(w_p + g(\delta_i, \theta_i)\right) e_i\right] - \frac{e_i^\gamma}{\gamma}$, in which the intrinsic and extrinsic returns also increase with effort at a declining rate, but optimal effort is independent of motivation levels: a marginal increase in motivation always increases utility by $\frac{1}{e}$, regardless of effort levels, which is implausible.

Naturally, if individuals find gaming incentives or the task to be unpleasant, the comparative statics switch signs.

The interaction of gaming incentives with other motivations is also of interest, since organizations frequently, and even inevitably, must combine them. First considering the interaction of extrinsic and intrinsic incentives, it is immediately clear that extrinsic incentives unambiguously suppress the effects of intrinsic motivation. For example, the interaction of extrinsic incentives (piece rate compensation) with gaming incentives is:

$\frac{\partial e_i^{*2}}{\partial w_p \partial \delta_i} = \frac{1}{\gamma}\left(\frac{1}{\gamma} - 1\right)\left(w_p + g(\delta_i, \theta_i)\right)^{\frac{1}{\gamma}-2} g_\delta$. If gaming incentives have a positive impact on utility (and therefore effort), then its interaction with extrinsic motivation, $\frac{\partial e_i^{*2}}{\partial w_p \partial \delta_i}$, is unambiguously negative (recalling that the value of effort - γ - is greater than one). The larger are extrinsic incentives, the smaller the effect of gaming incentives (and task motivation) on effort. This prediction is in line with models of crowding out (Frey and Oberholzer-Gee, 1997). If task or gaming motivation exert a negative effect on utility, the interaction of extrinsic utility is positive: higher extrinsic motivation offsets the negative effects on effort of gaming incentives.

Assuming task motivation and gaming incentives are positive, the interaction between extrinsic incentives and gaming incentives is unambiguously negative, since each diminishes the marginal effect of the other. In contrast, the interaction of the two intrinsic motivations, task and gaming, is potentially positive, since one may enhance the motivational effect of the other:

$$\frac{\partial e_i^{*2}}{\partial \theta_i \partial \delta_i} = \frac{1}{\gamma}\left(w_p + g(\delta_i, \theta_i)\right)^{\frac{1}{\gamma}-1} g_{\delta\theta} + \frac{1}{\gamma}\left(\frac{1}{\gamma} - 1\right)\left(w_p + g(\delta_i, \theta_i)\right)^{\frac{1}{\gamma}-2} g_\delta g_\theta \gtrless 0.$$

The first term is the direct effect of an increase in either task motivation or gaming. If these interact positively with each other, $g_{\delta\theta} > 0$, this term is positive. The second term is negative, capturing the diminishing returns to additional motivation, whether intrinsic or extrinsic. If diminishing returns are sufficiently small and the synergies between task motivation and gaming incentives are positive and large, the interaction of task and gaming yield greater effort.

Summing up, among individuals for whom both task motivation and gaming incentives increase utility, the effects on effort of task motivation, gaming incentives, and extrinsic motivation should exhibit a negative interaction. The effects on effort of task motivation and gaming incentives should exhibit a positive interaction if both enter utility positively and diminishing returns to motivation are sufficiently low, as when workers are paid an unconditional wage rather than given piece rate compensation. On the other hand, if task motivation and gaming incentives have few or negative synergies ($g_{\delta\theta}$ small or negative), the interaction of the two is unambiguously negative.

Appendix D: Additional tables and robustness checks

Table D.1: Summary Statistics and Balance Table, Treatments vs. Baseline

Dependent variable:	Task Motivation	Gender (1=Female)	Age (in years)	Ability in task	Competitiveness (5=High)	Risk preferences (5=Risk-seeking)	Income (5=Far above average)	Clarity of instructions
	I	II	III	IV	V	VI	VII	VIII
SUMMARY STATISTICS mean (std. dev.)	3.06 (0.82)	3.32 (0.57)	9.22 (2.29)	3.09 (0.95)	2.95 (0.89)	4.35 (0.82)	0.59 (0.49)	20.08 (2.64)
Gaming feature 1 (Points)	0.096	0.169	-1.166	1.254**	0.074	0.003	0.260	0.209
	(0.60)	(0.12)	(0.21)	(0.01)	(0.58)	(0.99)	(0.20)	(0.18)
Gaming feature 2 (Badges)	0.290	0.113	-1.537*	0.569	-0.041	-0.203	0.096	-0.012
	(0.13)	(0.29)	(0.10)	(0.25)	(0.74)	(0.31)	(0.65)	(0.94)
Gaming feature 3 (Leaderboard)	0.039	0.110	-1.093	0.402	-0.012	-0.170	0.213	-0.029
	(0.83)	(0.32)	(0.25)	(0.49)	(0.92)	(0.41)	(0.29)	(0.87)
Gaming feature 4 (Story/Theme)	0.085	0.152	-1.093	-0.230	-0.010	-0.006	0.163	-0.045
	(0.63)	(0.15)	(0.25)	(0.67)	(0.94)	(0.98)	(0.43)	(0.82)
All Gaming features	0.269	0.143	-0.995	0.420	-0.050	0.205	-0.033	-0.063
	(0.15)	(0.17)	(0.29)	(0.43)	(0.68)	(0.28)	(0.87)	(0.72)
Piece rate	0.210	0.171*	-0.711	0.611	-0.106	0.021	0.163	0.375***
	(0.27)	(0.09)	(0.45)	(0.19)	(0.36)	(0.91)	(0.38)	(0.01)
Piece rate + All gaming features	0.317	0.110	-1.643*	1.252**	0.006	0.330	0.063	0.046
	(0.11)	(0.32)	(0.08)	(0.01)	(0.96)	(0.12)	(0.76)	(0.80)
Constant	2.894***	0.465***	21.09***	8.698***	3.344***	3.070***	2.837***	4.279***
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R-Squared	0.019	0.011	0.031	0.044	0.008	0.029	0.012	0.035
P	0.446	0.804	0.101	0.017	0.922	0.146	0.759	0.005
Observations	362	362	362	362	362	362	362	362

Notes: OLS regressions with robust standard errors. The dependent variables are: task motivation (model I); gender (model II); age (model III); ability in coding task (model IV); competitive preferences (model V); risk preferences (model VI); relative income (model VII); and clarity of instructions (model VIII). * 10%, ** 5%, *** 1% significance level. P-values in parentheses.

Table D.2: Treatment effects on effort provision (Words coded correctly)

Dependent Variable: Effort (Words correctly coded)				
	I	II	III	IV
Gaming feature 1 (Points)	0.238	0.152	0.092	0.089
	(0.29)	(0.32)	(0.32)	(0.32)
Gaming feature 2 (Badges)	0.472	0.402	0.290	0.320
	(0.32)	(0.33)	(0.33)	(0.33)
Gaming feature 3 (Leaderboard)	0.491*	0.436	0.421*	0.458*
	(0.25)	(0.27)	(0.25)	(0.27)
Gaming feature 4 (Story/Theme)	0.846**	0.797*	0.737*	0.757*
	(0.42)	(0.42)	(0.42)	(0.43)
All Gaming features	1.031***	0.972***	0.865***	0.893***
	(0.35)	(0.34)	(0.33)	(0.33)
Piece rate	10.67***	10.61***	10.54***	10.45***
	(0.73)	(0.74)	(0.74)	(0.74)
Piece rate + All gaming features	10.37***	10.29***	10.18***	10.20***
	(0.76)	(0.76)	(0.73)	(0.74)
Gender		0.182	0.311	0.334
1 = Female		(0.31)	(0.33)	(0.33)
Age (in years)		-0.024	-0.023	-0.025
		(0.05)	(0.05)	(0.05)
Ability in task		0.022	-0.006	-0.034
		(0.05)	(0.05)	(0.06)
Task motivation			0.452**	0.423**
5 = Motivated			(0.18)	(0.18)
Competitiveness			0.410	0.351
5 = Highly competitive			(0.27)	(0.28)
Risk preferences				0.003
5 = Risk-seeking				(0.17)
Income (relative to others)				-0.093
5 = Far above average				(0.16)
Clarity of instructions				0.308*
5 = Always clear				(0.18)
Constant	0.884***	1.119	-1.398	-1.913
	(0.14)	(1.23)	(1.42)	(1.46)
Observations	362	362	362	362
Log likelihood	-810.8	-810.5	-806.4	-805.0
P	0.000	0.000	0.000	0.000
R-squared	0.210	0.211	0.215	0.216

Note: OLS specifications with robust standard errors in parentheses. * 10%, ** 5%, *** 1% significance level.

Table D.3: Treatment effects on effort provision (Words attempted)

Dependent Variable: Effort (Number of words attempted)	I	II	III	IV
Gaming feature 1 (Points)	0.238	0.152	0.092	0.089
	(0.29)	(0.32)	(0.32)	(0.32)
Gaming feature 2 (Badges)	0.472	0.402	0.290	0.320
	(0.32)	(0.33)	(0.33)	(0.33)
Gaming feature 3 (Leaderboard)	0.491*	0.436	0.421*	0.458*
	(0.25)	(0.27)	(0.25)	(0.27)
Gaming feature 4 (Story/Theme)	0.846**	0.797*	0.737*	0.757*
	(0.42)	(0.42)	(0.42)	(0.43)
All Gaming features	1.031***	0.972***	0.865***	0.893***
	(0.35)	(0.34)	(0.33)	(0.33)
Piece rate	10.67***	10.61***	10.54***	10.45***
	(0.73)	(0.74)	(0.74)	(0.74)
Piece rate + All gaming features	10.37***	10.29***	10.18***	10.20***
	(0.76)	(0.76)	(0.73)	(0.74)
Gender		0.182	0.311	0.334
1 = Female		(0.31)	(0.33)	(0.33)
Age (in years)		-0.024	-0.023	-0.025
		(0.05)	(0.05)	(0.05)
Ability in task		0.022	-0.006	-0.034
		(0.05)	(0.05)	(0.06)
Task motivation			0.452**	0.423**
5 = Motivated			(0.18)	(0.18)
Competitiveness			0.410	0.351
5 = Highly competitive			(0.27)	(0.28)
Risk preferences				0.003
5 = Risk-seeking				(0.17)
Income (relative to others)				-0.093
5 = Far above average				(0.16)
Clarity of instructions				0.308*
5 = Always clear				(0.18)
Constant	0.884***	1.119	-1.398	-1.913
	(0.14)	(1.23)	(1.42)	(1.46)
Observations	362	362	362	362
Log likelihood	-810.8	-810.5	-806.4	-805.0
P	0.000	0.000	0.000	0.000
R-squared	0.210	0.211	0.215	0.216

Note: OLS specifications with robust standard errors in parentheses. * 10%, ** 5%, *** 1% significance level.

Table D.4: Treatment and task motivation interactions (Words coded correctly)

Dependent Variable: Effort (Words correctly coded)	I	II
Gaming feature 1 (Points)	4.743	-3.119
	(7.58)	(9.65)
Gaming feature 2 (Badges)	6.341	7.165
	(5.90)	(7.18)
Gaming feature 3 (Leaderboard)	-0.541	-4.472
	(7.82)	(8.35)
Gaming feature 4 (Story/Theme)	18.39*	22.32**
	(9.62)	(11.28)
All Gaming features	-16.58**	-11.040
	(8.19)	(8.19)
Piece rate	125.5***	122.3***
	(20.05)	(19.80)
Piece rate + All gaming features	43.65*	40.81*
	(25.31)	(24.70)
Task motivation	4.305***	3.652**
5 = Motivated	(1.32)	(1.49)
Task motivation X	-0.533	1.162
Gaming feature 1 (Points)	(2.44)	(3.08)
Task motivation X	-0.695	-1.189
Gaming feature 2 (Badges)	(2.28)	(2.59)
Task motivation X	1.770	2.937
Gaming feature 3 (Leaderboard)	(2.82)	(2.97)
Task motivation X	-4.297	-5.531
Gaming feature 4 (Story/Theme)	(2.94)	(3.49)
Task motivation X	7.864**	5.796*
All Gaming features	(3.17)	(3.12)
Task motivation X	-8.946	-8.517
Piece rate	(6.42)	(6.35)
Task motivation X	16.15**	16.14**
Piece rate + All gaming features	(7.94)	(7.79)
Gender		1.008
1 = Female		(3.22)
Age (in years)		-0.219
		(0.43)
Ability in task		1.616***
		(0.61)
Competitiveness		0.903
5 = Highly competitive		(2.79)
Risk preferences		1.521
5 = Risk-seeking		(1.62)
Income (relative to others)		-0.607
5 = Far above average		(1.42)
Clarity of instructions		2.347
5 = Always clear		(1.77)

Constant	4.054	-19.970
	(3.51)	(14.16)
Observations	362	362
Log likelihood	-1697.0	-1690.6
P	0.000	0.000
R-squared	0.715	0.724

Note: OLS specifications with robust standard errors in parentheses. * 10%, ** 5%, *** 1% significance level.

Table D.5: Treatment and task motivation interactions (Words attempted)

Dependent Variable: Effort (Number of words attempted)	I	II
Gaming feature 1 (Points)	3.520	-4.992
	(7.73)	(9.88)
Gaming feature 2 (Badges)	3.584	4.424
	(6.18)	(7.19)
Gaming feature 3 (Leaderboard)	-4.512	-8.723
	(8.76)	(8.81)
Gaming feature 4 (Story/Theme)	21.28**	24.61**
	(9.07)	(10.69)
All Gaming features	-19.22**	-13.99*
	(8.63)	(8.48)
Piece rate	128.3***	125.2***
	(20.86)	(20.55)
Piece rate + All gaming features	45.96*	42.29*
	(25.51)	(25.12)
Task motivation	4.069***	3.386**
5 = Motivated	(1.46)	(1.57)
Task motivation X	-0.248	1.675
Gaming feature 1 (Points)	(2.53)	(3.19)
Task motivation X	0.012	-0.499
Gaming feature 2 (Badges)	(2.39)	(2.65)
Task motivation X	3.445	4.710
Gaming feature 3 (Leaderboard)	(3.22)	(3.23)
Task motivation X	-5.287*	-6.303*
Gaming feature 4 (Story/Theme)	(2.79)	(3.35)
Task motivation X	8.805***	6.802**
All Gaming features	(3.37)	(3.29)
Task motivation X	-8.774	-8.340
Piece rate	(6.67)	(6.59)
Task motivation X	16.46**	16.65**
Piece rate + All gaming features	(8.03)	(7.92)
Gender		0.491
1 = Female		(3.30)
Age (in years)		-0.282
		(0.46)
Ability in task		1.711***
		(0.64)
Competitiveness		0.654
5 = Highly competitive		(2.82)
Risk preferences		1.726
5 = Risk-seeking		(1.65)
Income (relative to others)		-0.778
5 = Far above average		(1.46)
Clarity of instructions		2.011
5 = Always clear		(1.84)

Constant	6.156	-14.920
	(3.84)	(14.69)
Observations	362	362
Log likelihood	-1708.7	-1702.5
P	0.000	0.000
R-squared	0.715	0.725

Note: OLS specifications with robust standard errors in parentheses. * 10%, ** 5%, *** 1% significance level.

Table D.6: Alternate measure of task motivation (robustness check – Rounds continued)

Dependent Variable: Quantity of effort provided (Rounds continued)	I	II	III
Gaming feature 1 (Points)	0.238	0.169	0.177
	(0.29)	(0.31)	(0.32)
Gaming feature 2 (Badges)	0.472	0.438	0.473
	(0.31)	(0.32)	(0.33)
Gaming feature 3 (Leaderboard)	0.491**	0.453*	0.509*
	(0.23)	(0.25)	(0.27)
Gaming feature 4 (Story/Theme)	0.845**	0.781*	0.806*
	(0.41)	(0.41)	(0.42)
All Gaming features	1.031***	1.001***	1.002***
	(0.33)	(0.33)	(0.33)
Piece rate	10.64***	10.63***	10.53***
	(0.72)	(0.73)	(0.73)
Piece rate + All gaming features	10.44***	10.41***	10.40***
	(0.75)	(0.74)	(0.75)
Task motivation	0.367***	0.357**	0.387***
5 = Motivated	(0.13)	(0.14)	(0.15)
Task motivation X	-0.010	-0.139	-0.030
Gaming feature 1 (Points)	(0.23)	(0.26)	(0.27)
Task motivation X	-0.006	0.018	-0.032
Gaming feature 2 (Badges)	(0.24)	(0.25)	(0.26)
Task motivation X	0.376	0.426*	0.305
Gaming feature 3 (Leaderboard)	(0.24)	(0.25)	(0.28)
Task motivation X	-0.489*	-0.480	-0.719**
Gaming feature 4 (Story/Theme)	(0.28)	(0.29)	(0.33)
Task motivation X	0.664**	0.647**	0.591*
All Gaming features	(0.30)	(0.32)	(0.31)
Task motivation X	-0.747	-0.713	-0.804
Piece rate	(0.80)	(0.79)	(0.79)
Task motivation X	1.397*	1.364	1.358
Piece rate + All gaming features	(0.82)	(0.84)	(0.84)
Gender		0.267	0.286
1 = Female		(0.33)	(0.33)
Age (in years)		-0.021	-0.024
		(0.05)	(0.05)
Ability in task		-0.022	-0.050
		(0.05)	(0.06)
Competitiveness		0.373	0.261
5 = Highly competitive		(0.27)	(0.27)
Risk preferences			0.067
5 = Risk-seeking			(0.17)
Income (relative to others)			-0.153
5 = Far above average			(0.15)
Clarity of instructions			0.321*
5 = Always clear			(0.18)

Constant	0.884***	0.134	-0.323
	(0.13)	(1.34)	(1.36)
Observations	362	362	362
Log likelihood	-802.2	-801.1	-799.2
P	0.000	0.000	0.000
Pseudo R-squared	0.219	0.220	0.222
Right censors	56	56	56

Note: Tobit specifications with robust standard errors in parentheses. * 10%, ** 5%, *** 1% significance level.