A Field Experiment in Motivating Employee Ideas

Michael Gibbs
University of Chicago & IZA

Susanne Neckermann
Erasmus University Rotterdam, ZEW & Tinbergen Institute

Christoph Siemroth
University of Mannheim

May 2014

We study the effects of a field experiment designed to motivate employee ideas at a large technology company. Employees were encouraged to submit ideas on process and product improvements via an online system. In the experiment, the company randomized 19 teams into treatment and control groups. Employees in treatment teams received rewards if their ideas were approved. Nothing changed for employees in control teams. Our main finding is that rewards substantially increased the quality of ideas submitted. Further, rewards increased participation in the suggestion system, but decreased the number of ideas per participating employee, with zero net effect on the total quantity of ideas. The broader participation base persisted even after the reward was discontinued, suggesting habituation. We find no evidence for motivational crowding out. Our findings suggest that rewards can improve innovation and creativity, and that there may be a tradeoff between the quantity and quality of ideas.

Keywords: creativity, incentives, innovation, idea suggestion systems, intrinsic motivation.

JEL classifications: C93, J24, M52, O32

We are very grateful to the unnamed company's executives and employees who provided us with access to their data, allowed us to interview them, and spent many hours answering questions about the data and company policies. We also thank participants at numerous seminars, as well as Jed DeVaro, Christoph Engel, Kathryn Ierulli, Sacha Kapoor, Mario Macis, Henry Sauermann, and Steve Levitt for their comments. Funding was provided by the Initiative on Global Markets at the University of Chicago Booth School of Business (Gibbs), the SEEK program of the state of Baden-Württemberg (Neckermann), and the DFG via SFB 884 (Siemroth).

1. INTRODUCTION

In today's economy, innovation is an important determinant of competitiveness and economic growth. "Ideation" – the generation and communication of novel and business-enhancing ideas – is therefore becoming increasingly important. One source of innovation that has traditionally received attention from academics and practitioners is R&D labs and initiatives, where activities are explicitly directed towards the generation of knowledge. A second source of ideas is the workforce more broadly. Employees in all hierarchical levels and organizational units may have useful knowledge about customers, and may have ideas about ways to improve processes, customer service, or product design. Outside R&D labs, job definitions typically do not specifically include innovation. Firms may therefore fail to capture this potential source of valuable ideas, small and large. While this is well recognized, ways to elicit knowledge and ideas from the workforce are not well understood.

One method to access employees' ideas is a formal suggestion system. Such systems are increasingly used across a wide set of industries. For example, Ohly et al. (2013) find that approximately 1 in 3 workplaces in the US and United Kingdom uses a suggestion system. Suggestion systems encourage employees to submit ideas for process and product improvements. At the same time, they make ideas transparent to management, who can then make decisions about implementation, and share the knowledge within the company. To date, little is known about how such systems work, and whether rewards tied to submitting ideas increase participation and ideation.

The aim of this paper is twofold. First, we provide a detailed picture of the ideation process in the study company, complementing the existing literature on employee suggestion systems. Second, we analyze the outcomes of a randomized field experiment designed to test the effectiveness of rewards for ideation, contributing to the debate among academics and practitioners about whether or not rewards improve or undermine innovation and creativity. In doing so, we also provide evidence on the relationship between employee characteristics and the quantity and quality of ideas suggested.

The setting of this study is a large, multi-national business process outsourcing company. It has a formal process by which employees submit new ideas, which are tracked through review and implementation. Data from this process present a unique opportunity to empirically study ideation. In an attempt to further improve the suggestion system, the company ran an experiment to test the effects of rewards for the submission and acceptance of new ideas. The company chose 19 established accounts, which we refer to as client teams, and randomly assigned the teams to treatment or control group. Overall, more than 11,000 employees participated in the experiment. For 13 months, employees in treatment teams received points for each idea submitted and approved. Additional points could be earned for favorable client feedback. Employees accumulated these points for redemption at an online site offering consumer goods. Apart from the reward program, policies were identical between treatment and control groups. Our data includes the treatment period as well as the 13 months leading up to the experiment. This allows us to use a difference-in-differences approach to estimate the effects of rewards on ideation, which accounts for possible pre-treatment differences between control and treatment group due to the relatively small number of client teams.

The experimental reward scheme had mixed effects on ideation.[1] Our main finding is that the quality of ideas (measured as the percentage of ideas accepted for implementation, or the percentage of ideas pitched to the client) was substantially increased by the reward. Thus, the treatment worked as intended, as only authors of accepted ideas received rewards. Further, rewards increased the fraction of employees who contribute in the suggestion system. Apparently, the rewards induced individuals to think more about ideas and to familiarize themselves with the suggestion system. This is in contrast to the notion that creativity is primarily a function of personality traits, rather than effort-based (George & Zhou 2001). However, the number of ideas submitted per potential contributor fell, suggesting that individuals put more work into the most promising ideas, which would explain the rise in idea quality. Combined,

[1] Throughout this paper we treat an increase in employee suggestions as positive for the firm, rather than a shift away from performance on core duties. In discussion with senior managers at the firm, we learned that they are not concerned about this potential multi-tasking problem. Rather, they hope that the reward program stimulates more employees to participate in innovation and continuous improvement, in addition to their daily duties.

there was no net effect on the total quantity of ideas submitted. We also find no evidence that the reward scheme crowded out intrinsic motivation (Deci 1971; Frey & Oberholzer-Gee 1997; Frey & Jegen 2001). Finally, we analyze the data from a 13-month window after the experimental reward scheme ended, and find some evidence of habituation effects.

There are large literatures in economics, psychology, social psychology, and management on innovation and creativity, too vast to survey here. Some research focuses on how personal characteristics or motivations affect creativity (Guilford 1950; Amabile 1983; Sauermann & Cohen 2010). Other researchers look at organizational variables such as perceived support, culture, or leadership (Amabile 1996; Robinson & Stern 1997; Tierney, Farmer & Graen 1999). Many studies look at the effect of rewards on creativity, though findings are mixed (Amabile 1982, 1996; Eisenberger & Armeli 1997; Deci et al. 1999; Joussemet & Koestner 1999; Eisenberger & Rhoades 2001; Eckartz et al. 2012). Some studies analyze how the structure of rewards (e.g., magnitude; short or long term) affects creativity (Ariely et al. 2009; Azoulay et al. 2011; Ederer & Manso 2013) or how rewards interact with the type of creativity: divergent or convergent (Charness & Grieco 2013). These studies provide valuable insights into how the originality of ideas is affected by rewards. However, they typically do not consider the submission process. This is important, since ideas need to be formulated and communicated to be effective. Furthermore, using a field setting is important for studying ideation as both the setting (experienced subjects working on a familiar task, where payments may be expected) and the task (innovation at work involves creativity, planning and execution) differ from those performed in a typical laboratory experiment. This study therefore complements existing studies on creativity in these respects.

This study is closely related to a small number of papers that look at employee suggestion programs. For the most part this literature uses survey data to elicit employees' motivations and organizational antecedents for submitting ideas. Leach et al. (2006) use data from 182 organizations in the UK and show that the number of ideas submitted correlates with scheme characteristics such as centralization, publicity, and use of rewards. Based on three case studies (KPN, Shell and Xerox), van Dijk and van den Ende (2002) argue that reward criteria affect the number of submissions. Ohly et al. (2013) provide a

conceptual discussion of suggestion schemes and an overview of the literature. Key indicators of the success of such systems are participation (percentage of employees who participate in the system), adoption (the percentage of submitted ideas that are implemented), and savings realized – indicators that we use in this study. Toubia (2006) models the effect of incentives on ideation, focusing on free-riding within groups, and concludes that incentives may be used to improve idea generation. He also provides evidence from a lab experiment consistent with that conclusion. Our paper builds on these prior studies by focusing on similar questions, using data from an actual employee suggestion system.

Although the form of the reward scheme might be specific to this setting, the results are of general interest. To our knowledge, this study is the first to offer causal evidence on the effect of rewards on ideation in the field. Furthermore, we are among the first to describe and analyze a process for encouraging employee innovation (the "Idea Portal") that is becoming more common as firms attempt to formalize innovation processes. In addition, we document a company's attempt to learn and improve its organizational design via field experimentation.

2. INSTITUTIONAL SETTING

Company background

The experiment took place at a large Asian information technology services company.[2] The company provides a variety of services, including outsourcing of business processes, R&D (including in some cases product design), and development of software or hardware solutions needed by clients. Most of the firm's clients are large global companies, without focus in a specific sector. The company is expanding and has well over 50,000 employees in many countries. The business process outsourcing industry has become increasingly competitive and commoditized. The company seeks to differentiate itself by offering greater innovation to clients, in the hopes of increasing client retention, growth, and profit margins. For that reason, all employees are encouraged to (in addition to their daily project work) suggest ideas for

[2] For confidentiality reasons, we cannot identify the company. Gibbs consulted with the company for 5 hours on an unrelated matter years ago. None of the authors received compensation for this project.

process or product improvements that are reviewed and, if approved, shared with the client. As another initiative designed to foster creativity, the company ran a field experiment to test the impact of material rewards on idea generation. It is these two programs – the idea generation system, and the additional experimental rewards program – that we study in this paper.

Because the firm provides business process outsourcing, most employees (and all on the teams studied here) reside in, and are citizens of, the company's native country. Employees are assigned to a specific client team. They do all of their work for this client. An employee is usually assigned to the same team for many years, so almost all workplace interactions take place within client teams. Among other things, this ensures confidentiality of sensitive client data. For our purposes this segmentation is useful, because client teams work independently of each other.

The "Idea Portal"

In an attempt to encourage employees to suggest new ideas, and to create a culture of innovation and collaboration, the firm set up an Idea Portal (our term). This is an intranet-based system designed to collect, evaluate, implement, and track new ideas. All employees, regardless of position or level, are encouraged to come up with new ideas and submit them, even if the idea is of small value. The process by which a new idea is handled is depicted in Figure 1:

[FIGURE 1 ABOUT HERE]

1. *Ideation*: one or more employees come up with a new idea that they deem worthy of suggesting. Employees may form ideation groups, and ideas can be submitted by 1 to 3 employees. An employee can be part of several ideation groups at the same time. Valid ideas might benefit a client directly or indirectly. Examples include new products or services, improvements to existing software systems, process improvements, or new software tools.

2. *Submission:* the employees submit the idea on the Idea Portal. Submission includes a brief description of the idea as well as its implementation, and estimates of implementation costs and projected revenue.

3. *Supervisor Input:* the supervisor is notified by the Portal as soon as an idea is submitted. This manager gives an initial review within 3 days. Often, the manager talks to the employee to refine or clarify the idea. This may be needed if the employee does not have the skills or experience to place the idea in the broader context of the business, estimate costs and benefits, etc.

4. *Senior Management Review:* the idea is evaluated by a panel of senior managers within the team's business unit, who meet regularly for this purpose. In some cases the idea is sent back for further refinement before the panel makes an accept/reject decision.

5. *Client Review:* if appropriate, approved ideas are brought to the client for final accept/reject decision. Clients may reject ideas, even if they have been approved internally, if they do not believe that the idea adds enough value given the costs of implementation.

6. *Implementation:* accepted ideas are implemented, and the results are tracked and evaluated. Where applicable, clients are asked to rate the quality of the idea on a scale of 1 to 4 points after implementation. Estimated costs and revenue may be updated based on client input about actual implementation.

All ideas input in the Portal are accessible by all employees, with the hope that this will spur ideation and spread ideas across the organization. However, to ensure client confidentiality, detailed information on posted ideas is visible only to employees within a client team, and to higher level managers across the company. The Idea Portal was a mature, robust system prior to the running of the experimental reward scheme. It had already processed several thousand ideas.

Experimental reward program

Incentives to suggest ideas were present in both the treatment and control groups prior to the experiment, stemming from intrinsic motivation, public recognition from ideas being formalized and possibly implemented, and career concerns. The formal procedure ensured recognition from clients, the team, the supervisor and reviewing managers, and potentially the rest of the organization. Until the experiment, however, there was no formal reward for contributing an idea.

Reflecting its culture of innovation, the company decided to run an experiment to see if rewards could improve ideation via the Portal. First, the company wanted to encourage more ideas that directly benefit the client, as compared to ideas that improve internal processes. This was intended to align employees with its strategy of developing closer working relationships with, and becoming an important source of innovation for, clients. Second, the company hoped to encourage employees to submit ideas on the Portal rather than implement them on their own. Doing so would increase transparency, allowing management to observe improvements, and decide whether or not to communicate them to clients. More-

over, the company hoped that ideas entered into the Portal would foster knowledge sharing and knowledge spillovers across the company.

The company has a bonus program that allows employees to earn points for good performance and other behaviors that the firm wants to motivate, such as contributions to a specific initiative, project completion, or job anniversaries. Employees accumulate points over time, similar to a frequent flyer program, and exchange them for consumer goods (e.g., a new smartphone) or shopping vouchers at an online store.[3] Such reward programs are not unusual in Asian companies. The bonus program applies to all employees and had been in place for several years prior to the experiment, just like the Idea Portal. For the experiment, the company decided to make use of this bonus program by offering rewards in the form of bonus points.

The experiment was run with 19 established client teams. Senior executives participated in the initial selection of these teams, ensuring that only teams with active use of the Idea Portal, and of a certain size, were used for the experiment.[4] Thus, participating teams were pre-selected from the entire workforce, and our analysis is conditional on this selection. However, selection for teams that use the portal regularly is arguably advantageous for our purposes, since it makes study teams more homogeneous and our focus is on the experimental reward. If anything, it should yield conservative estimates for the effect of rewards on ideation. Those teams were randomly assigned to treatment or control group. Overall, roughly 11,400 employees were involved in the experiment (6,000 in control teams, 5,400 in treatment teams). Employees in treatment teams received additional reward points for accepted ideas, while employees in control teams saw no change in their reward program.

The experimental reward was designed to motivate quality, and not just quantity. Submission of an idea alone did not merit a reward. If the idea was accepted for implementation, each member of the

[3] We were given an overview of how many points were redeemed, and for which goods, in July 2011. Gift vouchers were most frequently demanded, followed by goods from the category "Kids." Evidently, the points are often used for gifts to others.

[4] Teams used in the experiment met certain criteria for size and innovation activity: more than $2 million annual revenue, at least 50 employees, regular idea contribution (at least 5 ideas generated per month, with at least 1 shared with the client), and average client satisfaction index of at least 25 on a scale of –100 to 100. The company did not want to run the experiment with teams that were too small or not yet well-established.

ideating team earned 2,000 points. After implementation, authors could earn additional points depending on the client's rating. Unlike points for acceptance of the idea, these were divided equally among the ideators. To get a sense of the magnitude of the rewards, the 2,000 points awarded for idea acceptance were worth approximately 2.2% of monthly after-tax salary for lower level employees. If the idea received the highest client rating of 4, the employee earned an additional 40% of monthly net salary.[5] Therefore, rewards for implemented ideas were significant (in many firms, average annual bonuses may be not much more than 2.2%), and rewards for ideas rated highly by clients were very significant.

Apart from the change in the reward program, there were no policy differences between the treatment and control groups. Review processes were not changed in treatment teams and were, hence, identical to those in control teams. There were also no changes in supervisor or manager incentives in either treatment or control teams during the period. Finally, communications about innovation and ideation trainings were corporate-wide, and not affected by the experiment. While employees in treatment teams were informed about the reward scheme, they were not told that they were part of an experiment. Employees in control teams, managers on review panels, and clients did not know about the experiment. Only higher level management and human resource employees responsible for handling the reward program were aware of the experiment.

A concern in such settings may be spillover effects between control and treatment teams, or from HR to control teams. We believe that such concerns are likely to be limited in this case. Employees generally stay in their client team for many years.[6] Most or all workplace interactions are with team members and the client. Many teams are physically segregated; i.e., work at different company locations. Our discussions with HR managers indicated that the experimental program was not discussed outside of treatment teams, in part out of concerns that other teams would request the additional reward.

[5] Company management is convinced that the rewards were substantial and that employees were motivated by them. One employee told us that he took home an LCD TV because of the ideas he logged and implemented, and argued that this was a huge benefit given the price of the TV.

[6] In our data, 0.9% of employees switch client teams, and 0.1% switch from a control to a treatment team or vice versa. Since we have employee background data at only two points in time (prior to and during the experiment), these numbers refer to changes in team composition between those two points. According to management, no employee switched between treatment and control teams during the treatment period, so the 0.1% of switchers must have changed team prior to the experiment.

The company designed and conducted the experiment on its own. The executive in charge of the Portal and the experiment was a former student of one of us, so we had high-level access to the company to clarify questions. We visited the company's headquarters after the experiment had been completed, to interview executives about the company's organization, strategy, culture, and innovation processes. We also interviewed employees in some of the control and treatment teams about their work and innovation, without disclosing the experiment or that we would be studying data from the Portal or reward scheme. We followed up with numerous telephone discussions with company management to better understand their policies, the Portal, the organizational context at the time of the experiment, etc.

3. PREDICTIONS

In this section we provide a brief discussion of the predictions that guide the empirical work. Employees enter ideas into the Idea Portal even without the experimental reward. This may be due to intrinsic motivation, a sense of duty, desire for status and recognition from ideation, or implicit incentives such as merit raises or hope for promotion. Therefore, this paper studies the incremental effect on employee innovation from introducing explicit rewards. There are three relevant dimensions to ideation, given our data: employee participation (attempt to develop ideas, and enter them in the Portal); quantity of ideas suggested conditional on participation; and quality of ideas suggested.

Consider first *participation* in the Portal. Throughout this paper, "participation" means that an employee is both willing and able to submit ideas. It does not mean, however, that a participating employee always submits an idea. It is possible that the employee tries to come up with a suitable idea, but does not succeed, or has insufficient time to enter the idea in the Portal. A formal definition of participation, and an explanation of how we empirically estimate it, is given in the methods section. Willingness to engage in ideation likely depends on the rewards and costs from doing so. In this firm, employees are encouraged to enter ideas (in both control and treatment groups), but many still do not participate in ideation. Presumably this is because the cost of thinking about ideas or using the system, on top of normal work, outweighs the perceived benefit. The experimental incentive plan provides an additional benefit

from suggesting ideas that are accepted, all other factors held constant. This should increase the likelihood that an employee participates in ideation; that is, is motivated to look for ideas, and to enter an idea into the portal if she has one.

The reward scheme might also affect the *quality* of ideas suggested by employees. Psychologists argue that there may be a cognitive tradeoff between quality and quantity in ideation.[7] Employees might, for example, face a choice between "exploration" – pursuing new directions – and "exploitation" – pursuing familiar directions for incremental innovation (Robbins 1952; Gittins 1979). The former is less likely to generate new ideas, but such ideas may have higher expected value or quality. The latter may generate more ideas, but of lower quality. Idea quality likely depends on the time and thought that an employee spends elaborating and describing the idea, which could have been spent on new ideas instead. Moreover, since the experimental scheme rewarded quality, not quantity of ideas, we expect average quality of ideas to increase, particularly among *Prior Ideators* (employees who submitted ideas in the pre-treatment period). A possible countervailing effect might come from employees who were not prior ideators. The marginal non-ideating employee in the pre-treatment period may be less creative than those who did participate. If he decides to participate due to the treatment, average creativity among ideating employees decreases, which might lower average idea quality.

The effects of the reward on the *quantity* of ideas are similarly unclear. If participation increases, that will tend to increase quantity. However, given a quantity-quality tradeoff, average quantity of ideas suggested by a given employee (conditional on participation) might fall. The net effect is ambiguous.

It is often argued that *intrinsic motivation* is an important component of creativity, and that pay for performance "crowds out" intrinsic motivation. There are several rationales for this idea in the literature. One is the overjustification effect (Deci et al. 1999). In this view, extrinsic rewards are more salient than intrinsic motivation, and motivation therefore shifts to the external source when extrinsic rewards are

[7] The potential cognitive tradeoff between for quantity and quality of ideas suggests a multitask incentive problem (Holmstrom & Milgrom 1991). That is consistent with how the firm designed the experimental incentive plan, which rewarded high quality ideas (ideas that were accepted) rather than merely the number of ideas.

instituted. An alternative interpretation is signaling (Benabou & Tirole 2006). The idea is that employees who ideate in the absence of rewards do so because they like to engage in that activity. Doing so in the absence of rewards serves as a signal to themselves and/or others that they are intrinsically motivated, dedicated employees. The theory assumes that individuals do not have perfect knowledge about their motivations and use their actions to make inferences about their own character type. When rewards are present, ideation also becomes attractive to extrinsically motivated types, so that its value as a signal for intrinsic motivation decreases.

The crowding out view suggests that any incentive effect of rewards may be mitigated by a reduction in intrinsic motivation, with an unclear net effect. However, if there is motivational crowding out, the net effect should vary with the extent to which employees are intrinsically motivated. The data allow us to crudely group individuals by their degree of intrinsic motivation, and to test this. Consider employees in the pre-treatment period, without rewards. On average, intrinsic motivation among employees who ideate (prior ideators) should be higher than among those who do not. Therefore, crowding out effects in the treatment period should be more pronounced for the first group.

Crowding out theories do not clearly distinguish between participation in ideation, and the quantity and quality of ideas, so our analyses consider all three. However, participation is the concept that seems most closely aligned with the literature on crowding out. If there is crowding out of intrinsic motivation, the experimental reward should have a smaller (possibly negative) effect on the likelihood of participation for employees who previously participated, compared to those who did not. It might also have similar effects on the quantity or quality of ideas.

4. DATA & METHODS

Idea data were collected from the database behind the Idea Portal. Each idea record contains unique employee IDs of idea authors, dates of submission and latest update, client team ID, and a verbal description of the idea. The data also include estimates of the dollar cost of implementation and projected contribution to revenue. These estimates are provided initially by employees in consultation with their

managers, but may be updated later upon completion of implementation, with input from the client. Thus, we can construct estimated profit of the idea as revenue minus cost.

Idea review and management also takes place on the Portal, so the current status of each idea (e.g., "under review," "rejected," "under implementation") is recorded. Finally, the Portal records client ratings on a four point scale. However, for most ideas this rating is missing because not all suggestions were shared with clients, and only a few clients made use of this feedback option.

We have records for all ideas suggested during the 13 months prior to treatment ("period 1," May 2009 to May 2010), the 13 months of treatment ("period 2," June 2010 to June 2011), and the 13 months after treatment ("period 3," July 2011 to July 2012). Our data on whether the idea was accepted or rejected was last updated in January 2013.

The ideation data only contain records of employees who submitted ideas, so we used employment rosters to collect information about employees who did not submit ideas. These rosters were made available to us for two points in time, April 2010 and July 2011. We have no record of exactly when employees entered or exited the company, so we cannot reconstruct the exact team roster at each point in time. However, the two data points provide a good approximation for the pre-experimental and experimental periods.[8] This allows us to add inactive employees to our ideation data set and to look at the percentage of ideating employees.

Finally, we collected data on employee characteristics: gender, age, tenure at the company, client team ID, and salary group. There are 9 salary ranges, which correspond to hierarchical levels. Level 0 is entry level and Level 8 is the highest level in compensation and responsibility (executives). Table A in the appendix contains descriptions of all variables used in the analyses and reported in the tables.

[8] For the quantity regressions analyzing the number of ideas submitted per employee, we added all inactive employees from the April 2010 employment roster (end of pre-treatment period) to both periods, and employees from the July 2011 roster (end of treatment period) only to the treatment period. An allocation of the April 2010 employees only to the pre-treatment period resulted in very similar estimates. This confirms that the difference-in-differences approach addresses the concern that we do not have the exact employee roster, because the incompleteness of information affects both groups in the same way.

Summary statistics

Table 1 presents descriptive statistics for explanatory variables by treatment group in the three periods (pre-treatment, treatment, and post-treatment). The two groups are similar in terms of age, tenure, gender, salary, and share of prior ideators. The only statistically significant difference is that employees in treatment teams were slightly younger in the pre-treatment period. We therefore control for individual factors, including age, in all regressions below. The results are, however, not sensitive to the inclusion of these variables. As is common in technology firms, employees are relatively young with a mean age of about 30. Tenure is relatively short with a mean less than 3 years. Almost 80% of employees are male. The number of employees increased substantially from period 1 to period 2. This reflects company growth as well as the fact that the employee data for period 1 is less complete than for period 2, as we have only roster 1 (from the end of period 1) in the first period, whereas we can use both rosters to approximate the number of employees in the second period.

[TABLE 1 ABOUT HERE]

Table 2 displays descriptive statistics on outcome variables by period and group. Despite randomization, characteristics of control and treatment teams are somewhat different. Employees in control teams were more likely to suggest at least one idea than employees in treatment teams, in both periods. Further, the share of ideating employees dropped in both groups in the treatment period, with a stronger drop in the control group. These differences highlight the need to control for pre-treatment differences and time trends in our statistical analyses.[9] Ideas are typically submitted by 1 employee, with an average varying between 1 and 2 authors.

[TABLE 2 ABOUT HERE]

The last columns present descriptive statistics on three measures of idea quality. The first two indicate decisions made about each idea: accepted for implementation (“Imp”), and shared with the client

[9] Pre-treatment differences are not unexpected due to the relatively small number of client teams. As long as pre-treatment differences are not correlated with response to the treatment, they will not affect treatment effect estimates because the difference-in-differences approach controls for initial differences. We conducted robustness checks by including the average period 1 performance of the team as additional explanatory variable for period 2 performance. Direction and magnitude of the estimated treatment effects are similar to the estimates presented below.

("Shared"). Both measures are the result of a review process and are only reported for ideas with finished review ("|Fin"). These measures reflect idea quality, since better ideas have a higher chance of acceptance and are more likely to be shared with clients. The percentage of ideas accepted for implementation increased in both the treatment and control groups over time, with a stronger increase in the treatment group. The probability of sharing an idea with the client increased in the treatment group, but decreased in the control group.

A third quality measure is estimated profits (net value), the difference between the estimated revenue and cost of an idea. This measure varied substantially between ideas. Some ideas had very small projected financial impact, while one idea was estimated to improve revenue by $22 million.

Unfortunately, the last potential measure, the client idea rating on a 4-point scale, was given only for 17% of the ideas. Moreover, there appear to be serious selection issues. 85% of the ratings are 3 or 4 points. Only 1 out of 306 ratings received the lowest score of 1. This suggests that clients only reported ratings when they were happy with the idea, and used other channels to report dissatisfaction. We therefore do not use client ratings in our analysis.

Methods

Measuring and explaining the quantity of ideas

One of our main dependent variables is the count variable *Number of Ideas*, the total number of ideas submitted by an employee in either the pre-treatment or treatment period. The unit of observation for this analysis is therefore an employee-period. A large fraction of employees (91.4%) does not submit ideas, resulting in a large number of zeros for this variable. The prevalence of zeros disqualifies simple count data models such as Poisson or negative binomial (NB).[10]

[10] Poisson assumes equality of the conditional mean and conditional variance, which results in biased estimates for overdispersed data (conditional variance larger than mean). A negative binomial regression (type II) generalizes the Poisson model by assuming that the variance is equal to $\mathrm{Var}(y_i) = \mu_i(1 + \alpha\mu_i)$, where α is an additional parameter to be estimated. The NB model reduces to the Poisson for $\alpha = 0$. Estimating the NB model with our data, we find significant overdispersion, indicating that a Poisson regression is unsuitable.

We therefore analyze the data using a zero inflated negative binomial model (ZINB), which is a generalization of the NB, and used in particular for overdispersed data. ZINB is a mixture model, which explains the dependent count data variable jointly with a negative binomial process and a logit process. The logit process accounts for those zeros that are not explained by the negative binomial distribution ("excess zeros"). The intuition is that two processes are simultaneously at work in the data generating process. (1) The logit process models participation; that is, the employee's basic willingness or ability to submit an idea. If that condition is fulfilled, (2) the second process models the generation and submission of ideas, where the number of ideas is drawn from the negative binomial distribution. This implies that a zero observed in the data could be caused either by an employee not participating in ideation (condition (1) not fulfilled), or wanting to ideate but not having an idea in that period (condition (1) fulfilled, but the draw in (2) is zero).

Formally, let $f_1 := \Pr(No\ excess\ zero)$ be parameterized as logistic function and f_2(y) be the probability mass of count y in the negative binomial distribution. Then the mass of count y in the zero inflated negative binomial model is:

$$g(y) = (1 - f_1) + f_1 f_2(0), \text{ if } y = 0,$$

$$g(y) = f_1 f_2(y), \text{ if } y > 0.$$

The NB model is a special case where $f_1 = 1$. The ZINB model is well suited for our data, as the two processes allow us to effectively address overdispersion and also have a meaningful interpretation in our context.[11] As mentioned above, the approach allows us to model the submission decision hierarchically. The logit process provides information on whether or not an employee participated in ideation. If the answer is "no," the observed number of ideas is zero. Such excess zeros depict that an employee is not participating: they might be unmotivated to look for ideas, might shy away from using the Portal, or might work on a project with little or no room for ideation. If, on the other hand, the answer is "yes," we may

[11] Because we cluster standard errors, the objective functions of the models are pseudo-likelihoods, not true likelihoods of the data. Consequently, common tests to discriminate between models (likelihood ratio test for NB v. ZINB; Vuong test for ZINB v. zero inflated Poisson) are not applicable. Similar tests for clustered data are not yet common. However, taking the pseudo-likelihoods as true likelihoods, then Akaike's information criterion, likelihood ratio tests, and Vuong tests all favor the ZINB significantly over Poisson, NB, and zero inflated Poisson.

still observe zero ideas from this employee, in case they did not come up with an idea, even though they were actively looking them and willing to use the system. Positive counts are only observed when both processes are positive; i.e., the employee was motivated to ideate and also had an idea.

The latent class character of the zero inflated model allows us to disentangle these two different causes of non-ideation, whereas standard count data models or OLS do not. The logit and negative binomial processes of the ZINB are specified as follows:

(1) Logit: $\Pr(No\ excess\ zero_{it}|\boldsymbol{X}_{it}) = \exp(\lambda_{it}^{L})/(\exp(\lambda_{it}^{L}) + 1)$.

(2) NB: Negative binomial model with conditional mean $\exp(\lambda_{it}^{NB})$.

Estimating treatment effects

In all regressions, we use linear or nonlinear difference-in-differences estimations to infer the causal effect of the treatment on ideation. Difference-in-differences addresses several potential problems; for example, pre-treatment performance differences, possible time trends (a decrease in ideas over time, because low hanging fruit have already been picked), and approximation of the number of non-ideating employees per period. In order to estimate the treatment effect, we specify λ_{it}^{j} that appears in both equations (1) and (2) as follows:

$$\lambda_{it}^{j} = \beta_1^{j} Treatment_i \times Period\ 2_t + \beta_2^{j}\ Period\ 2_t + \boldsymbol{X}_{it}^{\prime j}\gamma + C_i^{j}, j = \{L, NB\},$$

where β_1 is the difference-in-differences treatment effect estimate. *Treatment* is the treatment group dummy, *Period 2* is the treatment period dummy, $\boldsymbol{X}_{it}$ is a vector of employee characteristics, and C_i is the client team fixed effect that accounts for unobservable time-invariant heterogeneity between the teams. The time invariant dummy *Treatment* is not estimated separately, because it is absorbed by the client team fixed effects (every client team is either in the treatment group or not). Index *i* denotes the author and *t* denotes the period (pre-treatment or treatment period). To control for time trends or time-specific shocks, we use time fixed effects in the form of pre-treatment and treatment period dummies. Note that the coefficients of the logit (L) and NB process are not constrained to be equal, which allows us to determine whether treatment affects the two processes differently.

Employees within a client team regularly interact, so the assumption of independence between observations (employees) is likely violated. We therefore cluster standard errors in all of our regressions at the client team level to allow for arbitrary correlation in error terms within a team.[12]

5. RESULTS

Who ideates?

Before we evaluate the experiment, we provide a brief overview of the correlates of ideation with respect to employee characteristics. Even though this is not a causal analysis, the question is of intrinsic interest, as it provides insights into the ideation process and lays a foundation for understanding the reward effects analyzed below. The results are displayed in Table 3. Our measures of ideation are the probability of submitting at least one idea within 13 months (one period), and the number of ideas submitted. In all regressions involving nonlinear models, we report average marginal effects (AME).

[TABLE 3 ABOUT HERE]

We control separately for age and company tenure. Age is negatively related to both outcome measures. On average, an additional year of age decreases the probability that an employee submits at least one idea by about 0.3 percentage points, and decreases the number of ideas per employee by about 0.006. These estimates suggest that older employees are less likely to contribute to the Idea Portal, and submit fewer ideas on average. An explanation may be that younger employees are more creative. Alternatively, the Portal might be more attractive to them, as they are more used to using new technology. Finally, they may be more aware of recent technological developments, such as new software programming techniques.

[12] The low number (19) of clusters is a concern, because the cluster and heteroskedasticity robust standard errors used here are only valid asymptotically. However, we have reason to expect that our standard errors are valid. Bertrand et al. (2004, Table VIII) show (for OLS) that clustered standard errors exhibit only small bias for 20 clusters. If there is no effect in the artificially generated data, then the null hypothesis rejection rate based on these standard errors is 5.8%, when it should be 5%. Block bootstrap, the primary alternative, rejects the null in 13% of cases (their Table V). We also estimated several linear models with the wild bootstrap procedure proposed by Cameron et al. (2008). They demonstrated in Monte Carlo simulations that their procedure does not over-reject. The t statistics from this procedure are virtually identical to those obtained with clustered standard errors in our linear models.

The effect of tenure is positive and statistically significant for both measures. One might have expected that employees with longer tenure have already suggested ideas that came to mind, and therefore have fewer new ideas. Our findings suggest the opposite. This may be because employees who work for the company longer have a deeper understanding of the business and client, which leads to more ideas.

Finally, we look at the effects of salary groups, which are proxies for hierarchical level. Senior management (salary groups 4 and above) is the reference category. Low-level groups 0 and 1 are pooled, because group 0 is too small for meaningful analysis on its own. The only statistically significant effect is that ideating low-level employees contributed about 0.18 fewer ideas within 13 months than active high-level employees. The mechanism behind this finding might be similar to the one for tenure. Both measures correlate with the employee's firm-specific experience, skills, and job match. Employees higher up in the hierarchy tend to have greater responsibility, control more resources, manage more subordinates, and have a higher-level understanding of the firm's and client's business and needs.

Treatment effects

Quantity of ideas

We now analyze the effects of the field experiment. As a first step, we examine whether the experimental reward affected the quantity of ideas suggested. As described above, the quantity of ideas submitted is modeled hierarchically with a logit process that captures an individual's participation, and a negative binomial process that describes the number of ideas submitted, given willingness to ideate. Table 4 displays OLS estimates and average marginal effects (AME) for the zero inflated negative binomial model (columns 1 and 2), explaining the number of ideas an employee suggests per period. Columns 3 and 4 display marginal effects for both the logit and the negative binomial process in isolation. The ZINB processes in columns 3-4 match best with the idea generation and submission process that we are studying, so we focus on those estimates.

[TABLE 4 ABOUT HERE]

Looking at the partial effects first, the treatment effect is significantly positive in the logit process, suggesting that rewards increased the share of employees motivated to participate – to think about

ideas for product or process improvements, and to submit them on the Portal. According to the average marginal effect, the rewards are estimated to increase the share of potential contributors by almost 18 percentage points, which is also economically significant. This accords with our prediction described above.

Interestingly, this positive effect is countered by a decrease in the number of ideas per author, as indicated by the negative binomial process. The result suggests that potential contributors submit about 0.26 fewer ideas during the 13 month treatment as a result of the reward program. The overall effect of rewards on the quantity of ideas is economically and statistically zero, which is consistent with Figure 2; the positive and negative partial effects cancel out. Consequently, ideation is spread over more employees, who concentrate on fewer ideas. This is consistent with our discussion about the potential tradeoff between the quantity and quality of ideas.

However, the zero overall effect may also be explained by crowding out: intrinsically motivated employees might have been negatively affected by the introduction of rewards, inducing a decrease in the number of ideas they submit. Since we have no direct measure of employees' intrinsic motivation and creativity, we use data from the pre-treatment period to categorize individuals as intrinsically motivated. The idea is that intrinsic motivation should, on average, be stronger among those that ideated in the pre-treatment period without rewards than among those that did not. We proxy intrinsic motivation with the dummy *Prior Ideator*, which captures whether or not an employee submitted at least one idea in the pre-treatment period, and is still in our sample in the treatment period. The second condition ensures that these employees have the opportunity to contribute again. About 8% of employees fall into this category in period 1. Columns (5) and (6) present the results of a model that allows the treatment effect to vary by whether or not the employee is a prior ideator. The models we estimate are identical to the main analyses presented above, except that:

$$\lambda_{it} = \beta_1 Treatment_i \times Period\ 2_t + \beta_2 Treatment_i \times Prior\ Ideator_i \times Period\ 2_t + \beta_3\ Period\ 2_t$$
$$+\beta_4 Prior\ Ideator_i \times Period\ 2_t + \beta_5 Treatment_i \times Prior\ Ideator_i + \boldsymbol{X}_{it}'\gamma + C_i,$$

for both the logit and NB process.[13] The treatment effect ($\beta_1 + \beta_2$) in the logit process, for example, is the change in the fraction of potential contributors among prior ideators in the treatment group, relative to that same change among prior ideators in the control group.

The results suggest that prior ideators in the treatment group are not more likely to participate (logit process). This is not surprising, as the logit process can be interpreted as describing whether or not an individual is actively looking for ideas and able to use the Idea Portal. The latter is necessarily true for the subgroup of employees who submitted ideas in the past. However, rewards appear to have induced employees who had not previously suggested ideas to enter the pool of potential contributors, increasing the fraction of potential contributors by about 6.5 percentage points.[14] In the negative binomial process, treatment reduced the average number of suggested ideas by 0.76 for all other (potentially contributing) employees, while the effect on prior ideators is not significantly different from zero. The overall effect on the amount of ideas submitted (not displayed) is -.021 ideas for prior ideators, and -.052 for employees who had not previously suggested ideas. The evidence does not support the crowding out hypothesis that rewards have detrimental effects for intrinsically motivated employees. Finally, the findings are consistent with a cognitive tradeoff between quantity and quality in ideation, or multitask incentives that rebalanced in favor of quality. To shed more light on these latter questions, we now investigate the effect of the experimental reward program on idea quality.

Quality of ideas

For quality analyses we focus on individual ideas rather than individual authors. Idea j may have several authors N_j, so we split each idea into N_j observations, each with the individual characteristics $\boldsymbol{X}_i$ of author i, where $i = 1, \ldots, N_j$. Hence, the unit of observation is the author-idea in all quality analyses. To ensure that single and multi-author ideas get the same weight, we weight each idea j by $1/N_j$. This proce-

[13] The average treatment effect on prior ideators is $ATE = \frac{1}{N}\sum_{i=1}^{N}[F(\beta_1 + \beta_2 + \boldsymbol{X}_i'\gamma) - F(\boldsymbol{X}_i'\gamma)]$, where F is the nonlinear function of the model (e.g., logit), $\boldsymbol{X}_i'\gamma$ is the effect of the remaining covariates, and i denotes the observation.

[14] One might expect the treatment effect in column (3) to be between the treatment effects of the two subgroups reported in column (5). However, since this is a nonlinear model, and we estimate separate time trends for each subgroup in (4), this need not be the case.

dure allows us to explain idea outcomes with both author- and idea-specific variables. We use three quality measures: whether the idea was shared with the client, whether it was implemented, and its net value.

Consider first the dummy variable *Shared*, which equals 1 if an idea was shared with the client, and zero otherwise. The review process is structured such that every idea is first reviewed internally. During internal review the panel decides whether or not to share the idea with the customer for feedback or approval. *Shared* is a meaningful measure of quality for two reasons. First, it would be unwise of the company to bother clients with bad or trivial ideas. Second, the aim of the company was to increase its value-added as perceived by the customer. Value-added is more salient for ideas that are explicitly communicated to the client than for ideas that are implemented as by-products of ongoing business without communication to clients. In Table 5 we report both OLS coefficients and logit marginal effects of the following models:

OLS: $Shared_{ijm} = \mu_{ijm} + \varepsilon_{ijm}$ if $Finished_j = 1$,

Logit: $\text{Prob}\left(Shared_{ijm} = 1 \middle| \boldsymbol{X}_{im}, \boldsymbol{I}_j, Finished_j = 1\right) = \frac{\exp(\mu_{ijm})}{\exp(\mu_{ijm})+1}$,

where $\mu_{ijm} = \beta_1 \, Treatment_i \times Period\,2_m + \boldsymbol{X}'_{im}\gamma + \boldsymbol{I}'_j\delta + C_i + T_m$. *Treatment* represents the treatment group dummy, $\boldsymbol{X}_{im}$ is a vector of employee characteristics at the time of submission, $\boldsymbol{I}_j$ is a vector of idea-specific variables including project type, C_i is the client team fixed effect of employee *i*, and T_m is the monthly time fixed effect.[15] Index *i* denotes the author, *j* the idea, and *m* the month of idea submission. The results are robust to different ways of estimating the time trend (period or month fixed effects). The analysis only includes ideas with finished review, because information on our outcome variables *Shared* and *Implemented* are only available after review.

[TABLE 5 ABOUT HERE]

[15] A priori, one might have thought that the treatment affects the number of authors submitting an idea, which in turn might influence the probability that the idea is shared. The inclusion of the number of authors as a control variable would then bias treatment effect estimate. We looked into this and found no statistically or economically significant effect of the treatment on the number of authors (results available upon request). Therefore, we include *number of authors* as a covariate in all quality regressions to improve fit. The treatment effect estimates are even larger when *number of authors* is omitted as explanatory variable.

Columns (1) and (2) of Table 5 report results for the probability that an idea is shared with the client. We find a substantial and statistically significant positive treatment effect. Depending on the model, the treatment increased the probability of sharing the idea by 19 to 21 percentage points. This suggests that the treatment worked as intended, and employees in treatment teams focused on the quality rather than the quantity of ideas. This could be due to employees increasing the fraction of customer-related ideas (a shift in focus of ideas), or to employees increasing the quality of the submitted idea without such a re-shift in focus. Both should increase the likelihood that an idea is shared with the customer.

A second measure of quality is *Implemented*, which equals 1 if the idea was accepted for implementation, and 0 if not. *Implemented* is a measure of idea quality because the company has economic incentives to only approve worthwhile ideas. The difference from *Shared* is that an idea may be implemented without being shared with the client. Alternatively, ideas may be approved internally, but then rejected by the client (i.e., shared, but not implemented).

The models we estimate are the same as above for *Shared*, except that we now use *Implemented* as the dependent variable. The results are displayed in columns (3) and (4) of Table 5. Once more we find a large positive and statistically significant treatment effect. Rewards improve the likelihood of idea implementation by 15 to 18 percentage points.

When entering ideas in the Portal, authors are required to estimate the costs and benefits (in US dollars). Supervisors and clients may later refine those estimates based on their experience and knowledge. As a final quality proxy, we look at the expected profit from an idea, measured as the log of projected net value (benefit minus cost). One would think that the company only accepts ideas with positive net value. However, this is not the case. Strategic considerations, such as the development of new competencies, effects on client relationships, or synergies with other ideas induce the company to accept some ideas with a negative net value and reject some with a positive value. In our dataset about 25% of ideas with negative net value were accepted. This share increases to about 40% when we only consider ideas with finished review. Note that ideas with zero or negative net value drop out when taking the log-

transformation. We therefore also ran a linear model. It yielded the same qualitative results, but with inferior fit.

Table 5 column (5) shows OLS results for projected profit. The point estimates of the treatment effect are substantial and positive. A treatment effect estimate of 0.308 in the log-linear specification translates into an effect of about +36%. One interpretation is that employees knew that more profitable ideas are more likely to be accepted, so they shifted their focus to more profitable ideas when rewards were introduced. However, despite the large point estimate, the effect is not significantly different from zero, so we cannot reject the hypothesis that treatment had no effect on projected profit.

We also tested for motivational crowding out effects using all three quality measures, by comparing treatment effects on prior ideators with those on all other employees (using the same methodology as explained above in the quantity section). None of the regressions showed a significant difference in the treatment effect (not shown). Hence, there is no evidence of crowding out with respect to idea quality.

To sum up, we find large positive and statistically significant treatment effects on idea quality as measured by the number of ideas that are implemented and that are shared with the client, but cannot reject the null that projected profits are not affected by the experiment. In line with our findings on quantity, there is no evidence of crowding out on all measures of idea quality. The increase in idea quality might explain the drop in the number of ideas per author that we detected above. This provides support for the notion that ideation is subject to a quantity-quality tradeoff, with authors in treatment teams focusing on fewer but better quality ideas.

The regressions in Table 5 also include author characteristics. There are two things worth mentioning. First, tenure is positively correlated with the probability that an idea is shared with the client, as well as with net value. Given that these regressions control for salary group, this is in line with our finding of a positive correlation between tenure and the number of ideas suggested. Both results suggest that firm- and/or client-specific human capital may be valuable for ideation. Second, ideas with more authors have higher average quality, for all quality measures. This is consistent with research that shows that collaboration, particularly between individuals with different perspectives and skills, improves the quality of

innovation (Ford & Sullivan 2004; Shin & Zhou 2007). An alternative explanation is that authors pool ideas and submit the best idea among them. To illustrate, suppose *N* authors randomly form a group, and each author *i* has one idea with quality X_i. Group quality is the Nth order statistic $X_{(N)} = \max\{X_1, \ldots, X_N\}$, which is increasing in *N* even without formal collaboration.

Post-treatment effects

The company discontinued the reward scheme after the 13 months of the field experiment. We collected data from the Portal for the 13 months that followed. This allows us to examine whether the positive treatment effects on participation and idea quality persisted, which might suggest habituation, or disappeared. Alternatively, one might expect a reduction in idea quantity and quality – either because the incentive is removed, or due to perceived unfairness associated with removing the reward. Table 6 displays results on quantity and quality, now including a post-treatment effect (DID Post-treatment) for the 13 months following the experiment.

[TABLE 6 ABOUT HERE]

The post-treatment effect on the share of potential contributors (logit process of the ZINB model) remains positive and significant. Employees who became potential contributors due to rewards tend to continue participation after the reward is discontinued. This might reflect habituation. Another potential explanation is that rewards changed the working culture of treatment teams, with greater attention to, or peer pressure for, ideation. Alternatively, the reward might have better communicated the importance to the firm of employee ideas, thereby changing implicit incentives. Unfortunately, the data do not allow us to disentangle these alternative interpretations. The effect on the number of ideas per potential contributor (negative binomial process, column 2) is insignificant in the post-treatment period. The overall effect on the quantity of ideas remains statistically and economically insignificant.

The table also reports the results for one of our quality measures, *Shared* (effects are similar when using *Implemented* as the dependent variable). Interestingly, the treatment effect on quality vanished after elimination of the reward scheme. Hence, while the treatment effect on participation remains, the effect on quality disappears. Apparently, employees continue to think about and submit ideas, but they no longer

focus their effort on ideas with a high probability of acceptance. Instead they return to baseline with respect to idea quality. This could also explain why the effect on the number of ideas per participating employee becomes insignificant (column 2). Taken together, these findings are consistent with changes in multitask incentives during and after the treatment. In this view, the rewards induced employees to spend additional effort fine-tuning and improving ideas to increase their odds of receiving a reward. Once, the rewards were eliminated, relative incentives to emphasize quality returned to their original level. This is interesting for both theory and application, as it suggests that the quality of ideas can be manipulated with rewards.

Financial benefits of the experimental reward

We now provide back-of-the-envelope calculations of the profitability of the experimental reward program. The regression estimates indicate that the treatment effect on the quantity of ideas submitted was essentially zero. This implies that the program created little additional cost associated with evaluating and reviewing submitted ideas (opportunity costs of manager or employee time). Further, there were few costs associated with instituting the reward system, as the system of collecting and redeeming points was already in place, as was the Idea Portal.

The direct reward cost can be calculated with reasonable precision. 247 employees in the treatment group had an idea accepted in the 13 months of the treatment, and each received 2000 points for it. These idea acceptance points cost the company approximately $1,990. As mentioned before, the company also awarded substantial points for favorable client ratings. These amounted to about $14,250. Therefore, the total bonus cost of the experiment was approximately $16,240.

The benefits of the program cannot be estimated with precision, but must be inferred from estimates of the profitability of implemented ideas. Note that the treatment effect on the likelihood of implementation in Table 5 includes an indirect effect, via the effect of rewards on the net value of ideas. That might, in turn, affect the likelihood of implementation. To separate out the direct effect, we re-ran the regressions in Table 5 but controlling for net value. This results in a direct treatment effect on the likelihood of implementation of 15.2 percentage points. The point estimate of rewards on net value, as dis-

played in Table 5, is 36%. Taken together, these imply a combined increase in net value of about $8.3 million. If we use a linear instead of the semi-log specification in Table 5 to estimate the treatment effect on net value, the benefits amount to about $7 million.

One possible is the large variance in projected net value of the projects. We therefore re-estimated the treatment effect removing outliers (net value above $100k or below -$100k). This reduces the estimate of the treatment effect on net value from +36% to +26.7%. Excluding outliers from this computation and using this more conservative estimate, the financial benefit of the experiment still amounts to $1.5 or $1.4 million for semi-log or linear estimates respectively. Thus, for reward costs in the ten-thousands, the company was able to stimulate new ideas that generated financial value in the millions.

This analysis excludes some relevant costs and benefits that we cannot estimate with the available data. For example, more ideas were shared with clients under the reward scheme. This could imply benefits to the company in terms of higher perceived value-added, customer satisfaction, and loyalty. It also entails some risks, if an idea turns out to be a mistake. Ideas may create organizational benefits, such as knowledge spillovers across the organizations, development of new technology or competencies, or follow-up innovations based on earlier ideas.

6. CONCLUSIONS

This paper provides a unique statistical analysis of employee innovation. Data from a company's employee suggestion system allow us to analyze the effectiveness of rewards for fostering employee ideation, which were implemented as a randomized field experiment. Our findings are of interest for the debate on the causal effects of incentives on ideation, creativity, and intrinsic motivation. Most of the previous literature uses survey data and case studies to analyze ideation, and laboratory experiments to study creativity. This study complements the existing literature by studying experienced individuals performing their regular work. Moreover, random assignment to control and treatment groups, and use of difference-in-differences estimation, facilitate causal inferences.

Our findings suggest that rewards are a suitable tool to induce employees to think about process and product improvements, and to induce them to use a formal ideation system. The reward scheme substantially increased the likelihood that employees participate in the Idea Portal. Broadening the participation base was, in fact, one of the company's objectives. The data from the post-experimental period show that once familiar with the system, employees continued using the ideation system even in the absence of the reward scheme, suggesting habituation.

The firm's experimental incentive plan emphasized quality, since points were awarded only for ideas that were accepted for implementation. A main finding of this paper is that the treatment was successful in increasing idea quality, as measured by the percentage of ideas implemented. One interpretation is that reward criteria signal employees about the types of ideas that the company values. Interestingly, employees returned to baseline with respect to the number and quality of ideas they suggest when the rewards were discontinued. Taken together, the treatment and post-treatment responses on quality suggest that employees can fine-tune ideation and are responsive to incentives when doing so. The results are also consistent with the notion of a quantity-quality tradeoff, which may be addressed with multitask incentives. Proper choice of a performance measure is likely to be important. If the company had rewarded each submission rather than idea acceptance, we would likely have observed an increase in the quantity of ideas at the expense of quality.

Employees in this firm suggested ideas even without formal rewards, indicating intrinsic motivation or implicit incentives. However, the results do not provide support for the view that rewards crowd out intrinsic motivation. Not only did the reward scheme appear to motivate more employees to become active ideators, and to focus on idea quality, but employees ideating prior to the experiment did not reduce performance when offered the rewards.

The findings also illustrate the utility to firms from experimentation in organizational policies. The firm was able to test a new incentive before rolling it out to the entire organization. This allowed it to test the efficacy of the general idea ("do incentives enhance employee innovation?") by estimating its effects on various dimensions of employee performance. In addition to providing useful information

about whether the idea was sound, the experiment provided insights into the specific design of the policy. Conceivably the firm could have gone even further. For example, they might have tried different rewards, perhaps awarding points merely for suggesting ideas, or for suggesting ideas as a group of employees. There may be substantial costs to implementing organizational policies that are poorly designed, and bureaucratic inertia might make it costly to change or eliminate a policy after it is implemented. Moreover, changes to an existing policy might reduce management's credibility with employees. Experimentation can help firms reduce these risks from new or changed policies.

Our analysis has, of course, only scratched the surface on questions about employee innovation, and interactions between intrinsic and extrinsic motivation. Our findings suggest a tradeoff between quantity and quality in ideation. How strong is this tradeoff? What are the implications of this tradeoff for job design, team structure, and incentive plan design? Data on employee tasks and time management might allow deeper understanding of this issue. Nevertheless our findings are encouraging for the view that incentives do not undermine employee creativity if they are designed appropriately.

The company discontinued the experiment after 13 months of treatment. Its own analysis – a comparison of simple averages between treatment and control groups – suggested that the incentive was ineffective, and it decided not to implement the reward scheme company-wide. There are three main reasons for why they reached a different conclusion. First, our difference-in-differences approach controls for initial differences between the two groups. Second, we look at the subset of ideas that finished the review process, while the company looked at all submitted ideas. That might have introduced a confounding factor if review speeds varied (teams with faster reviews would have misleadingly-higher acceptance rates and vice versa). Third, we control for team composition, such as age, hierarchical level and gender. Differences between groups in those characteristics may affect conversion rates, but are not easily disentangled from the treatment effect using non-regression methods. After seeing our analysis, the company is now designing a similar reward scheme to roll out for the entire company.

REFERENCES

Amabile, T. M. (1982). Children's Artistic Creativity Detrimental Effects of Competition in a Field Setting. *Personality and Social Psychology Bulletin*, 8(3), 573-578.

Amabile, T. M. (1983). *The Social Psychology of Creativity*. New York: Springer-Verlag.

Amabile, T. M. (1996). *Creativity in context: Update to "the social psychology of creativity."* Westview Press, New York.

Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009). Large stakes and big mistakes. *The Review of Economic Studies*, 76(2), 451-469.

Azoulay, P., Graff Zivin, J. S., & Manso, G. (2011). Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics*, 42(3), 527-554.

Bénabou, R. & Tirole, J. 2006. "Incentives and Prosocial behavior". *American Economic Review*. N. 96, pp. 1652-1678.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1), 249-275.

Cameron, A. C., Gelbach, J., & Miller, D. (2008) "Bootstrap-Based Improvements for Inference with Clustered Errors", *Review of Economics and Statistics*, Vol. 90(3): 414-427.

Charness, G., & Grieco, D. (2013). Individual creativity, ex-ante goals and financial incentives. Working Paper.

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

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

Eckartz, K., Kirchkamp, O., & Schunk, D. (2012). How do incentives affect creativity? Working Paper.

Ederer, F., & Manso, G. (2013). Is Pay for Performance Detrimental to Innovation?. *Management Science*, 59(7), 1496-1513.

Eisenberger, R., & Armeli, S. (1997). Can salient reward increase creative performance without reducing intrinsic creative interest? *Journal of personality and social psychology*, 72(3), 652.

Eisenberger, R., & Rhoades, L. (2001). Incremental effects of reward on creativity. *Journal of personality and social psychology*, 81(4), 728.

Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys*, 15(5), 589–611.

Frey, B. S., & Oberholzer-Gee, F. (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *American Economic Review,* 87 (4), 746–755.

Ford, C., & Sullivan, D. M. (2004). A time for everything: How the timing of novel contributions influences project team outcomes. *Journal of Organizational Behavior*, 25(2), 279-292.

George, J. M., & Zhou, J. (2001). When openness to experience and conscientiousness are related to creative behavior: an interactional approach. *Journal of Applied Psychology*, 86(3), 513-524.

Gittins, J. C. (1979). Bandit Processes and Dynamic Allocation Indices. *Journal of the Royal Statistical Society, Series B (Methodological)*, 41(2): 148–177.

Guilford, J. P. (1950). Creativity. *American Psychologist,* 5, 444–454.

Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org.*, 7, 24.

Joussemet, M., & Koestner, R. (1999). Effect of expected rewards on children's creativity. *Creativity Research Journal*, 12(4), 231-239.

Leach, D. J., Stride, C. B., & Wood, S. J. (2006). The effectiveness of idea capture schemes. *International Journal of Innovation Management,* 10, 1-26.

Ohly, S., Leach, D. & Zou, Q. (2013). Suggestion-Making in Organizations: A Review and Multilevel Model. Unpublished Manuscript.

Robinson, A. G., Stern, S., & Stern, S. (1997). *Corporate creativity: How innovation and improvement actually happen*. San Francisco: Berrett-Koehler Publishers.

Robbins, H. (1952), Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5), 527–535.

Sauermann, H., & Cohen, W. M. (2010). What makes them tick? Employee motives and firm innovation. *Management Science*, 56(12), 2134-2153.

Shin, S. J., & Zhou, J. (2007). When is educational specialization heterogeneity related to creativity in research and development teams? Transformational leadership as a moderator. *Journal of Applied Psychology*, 92(6), 1709.

Tierney, P., Farmer, S. M., & Graen, G. B. (1999). An examination of leadership and employee creativity: The relevance of traits and relationships. *Personnel Psychology*, 52(3), 591-620.

Toubia, O. (2006). Idea generation, creativity, and incentives. *Marketing Science*, 25(5), 411-425.

van Dijk, C., & van den Ende, J. (2002). Suggestion systems: Transferring employee creativity into practicable ideas. *R&D Management,* 32, 387-395.

Figure 1. Process for Evaluating New Ideas

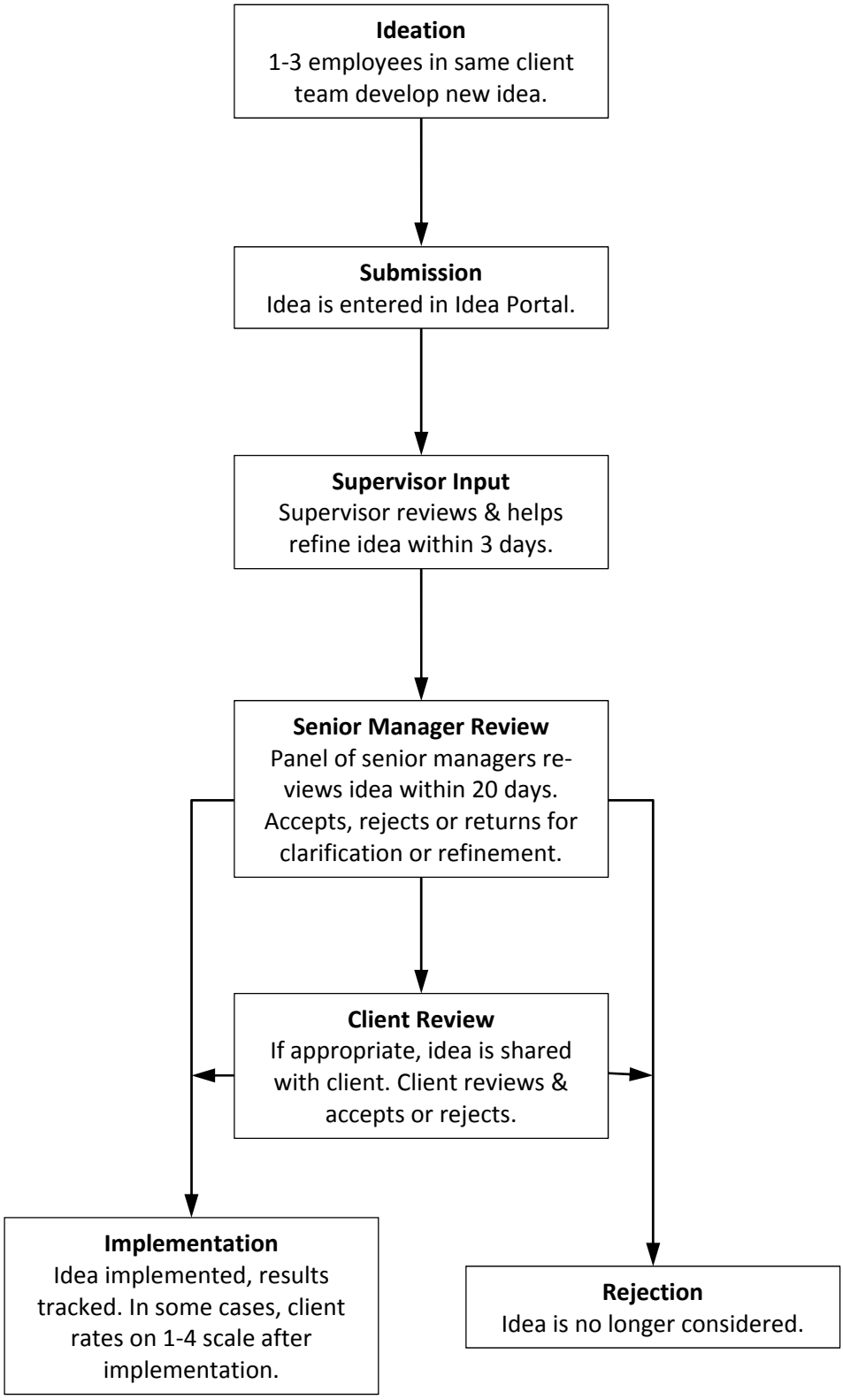

Table 1: Descriptive statistics

	Period	Treatment Group	Control Group	Combined
Number of Employees	1	3015	3185	6200
	2	5260	5881	11141
	3	4061	4511	8572
Mean Age	1	29.4**	29.9	29.7
		(4.70)	(4.87)	(4.80)
	2	29.7	29.6	29.7
		(5.38)	(5.41)	(5.40)
	3	30.5	30.3	30.4
		(5.50)	(5.54)	(5.52)
Mean Tenure	1	2.99	2.90	2.94
		(2.32)	(2.34)	(2.33)
	2	2.97	2.80	2.88
		(2.48)	(2.37)	(2.42)
	3	3.73	3.53	3.62
		(2.53)	(2.30)	(2.41)
Share of Men	1	0.78	0.80	0.79
		(0.42)	(0.40)	(0.41)
	2	0.76	0.77	0.76
		(0.43)	(0.42)	(0.43)
	3	0.76	0.77	0.76
		(0.43)	(0.42)	(0.43)
Mean Salary Group	1	1.43	1.45	1.44
		(0.73)	(0.77)	(0.75)
	2	1.47	1.47	1.47
		(0.79)	(0.82)	(0.81)
	3	1.49	1.49	1.49
		(0.81)	(0.84)	(0.82)
Share of Prior Ideators	1	0.087	0.063	0.075
		(0.28)	(0.24)	(0.26)
	2	0.049	0.035	0.042
		(0.22)	(0.18)	(0.20)
	3	0.064	0.045	0.054
		(0.24)	(0.21)	(0.23)

Note: Standard deviations of the means are displayed in parentheses. In each line, the difference of group means is tested with a t-test using standard errors that are clustered at client team level. ***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level. *Number of employees* in period 1 (pre-treatment) is based on employment roster 1; *Number of employees* in period 2 (treatment) is based on rosters 1 and 2; *Number of employees* in period 3 (post treatment) is based on roster 2. *Age* and *tenure* are measured at the end of the respective period.

Table 2: Summary statistics on outcome variables by group for pre-treatment and treatment period

Group	Period	# of ideas	Ideator*	Authors	Finished	Imp|Fin*	Shared|Fin*	Log(Net Value)
Treatment	1	517	0.195	1.716	0.745	0.410	0.449	8.705
			(0.355)	(0.893)	(0.436)	(0.379)	(0.440)	(2.030)
	2	566	0.083	1.309	0.643	0.516	0.597	9.428
			(0.264)	(0.550)	(0.480)	(0.461)	(0.450)	(2.023)
Control	1	361	0.336	1.903	0.634	0.638	0.415	8.865
			(0.285)	(1.118)	(0.482)	(0.249)	(0.398)	(2.083)
	2	363	0.112	1.402	0.510	0.712	0.339	9.257
			(0.204)	(0.667)	(0.501)	(0.366)	(0.446)	(2.146)

Note: Ideator is the share of employees who submitted at least one idea in the respective period. *Authors* is the mean number of authors per idea. *Finished* is the share of ideas with finished review; those ideas are either accepted for implementation or rejected. *Imp|Fin* and *Shared|Fin* denote the shares of ideas selected for implementation or for sharing with the client, respectively, among ideas with finished review. Log(*Net Value*) is the mean of the logarithm of the projected profit in US dollar terms. Period 1 is pre-treatment, period 2 is the treatment period. * denotes client team means (pooled standard deviations).

Table 3: Who ideates? Influence of employee characteristics on ideation

	(1) Logit AME	(2) ZINB AME
Dependent variable	Ideator	Number of Ideas
Age	-0.007***	-0.017***
	(0.001)	(0.004)
Tenure	0.018***	0.030***
	(0.003)	(0.009)
Male	0.004	-0.056
	(0.009)	(0.042)
Salary Groups 0 & 1 pooled	-0.094**	-0.516*
	(0.039)	(0.291)
Salary Group 2	0.005	-0.119
	(0.034)	(0.100)
Salary Group 3	0.075	0.022
	(0.048)	(0.074)
Client FE	yes	yes
Log Pseudo likelihood	-1416.53	-2177.45
Clusters	18	19
Observations	5887	5916

Note: The regression uses data from the pre-treatment period only, where both groups have identical incentives. *Ideator* is a dummy indicating whether an employee submitted at least one idea in the given period. *Number of Ideas* is the number of ideas submitted within the 13 pre-treatment months. *Salary group* is an indicator for an employee's position in the company hierarchy. The reference category is (upper) management, that is, salary groups 4 and above. The marginal effects of *Age* and *Tenure* are based on linear and quadratic terms. Standard errors are clustered at the client team level. ***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table 4: Amount of suggested ideas per employee and period

	(1) Overall effect OLS	(2) Overall effect ZINB AME	(3) Zero Inflated NB Logit AME	(4) Zero Inflated NB Negative Binomial AME	(5) Zero Inflated NB Logit AME	(6) Zero Inflated NB Negative Binomial AME
Dependent variable	NumIdeas	NumIdeas	Pr(NonExcessZero)	NumIdeas w/o ExcessZeros	Pr(NonExcessZero)	NumIdeas w/o ExcessZeros
DID Treatment	0.001	-0.004	0.178***	-0.264**	0.065***	-0.761***
	(0.063)	(0.051)	(0.049)	(0.132)	(0.019)	(0.281)
DID Treatment Prior Ideator					0.003	-0.207
					(0.056)	(0.331)
Age	-0.007	-0.005**	-0.008***	-0.004	-0.003**	0.014
	(0.008)	(0.002)	(0.002)	(0.008)	(0.001)	(0.010)
Age^2	-0.000					
	(0.000)					
Tenure	0.061***	0.027***	0.060***	0.005	0.018***	0.005
	(0.012)	(0.004)	(0.010)	(0.030)	(0.004)	(0.029)
$Tenure^2$	-0.004**					
	(0.002)					
Male	-0.016	-0.015	0.052***	-0.165*	0.013	-0.303**
	(0.014)	(0.015)	(0.013)	(0.086)	(0.009)	(0.141)
Controls salary groups	yes	yes	yes	yes	yes	yes
Client FE	yes	yes	yes	yes	yes	yes
R^2	0.088					
Log Pseudo likelihood		-5266.22	-5266.22	-5266.22	-4558.93	-4558.93
Clusters	19	19	19	19	19	19
Observations	17045	17045	17045	17045	17045	17045

Note: The table reports the results from OLS and zero inflated negative binomial models used to explain the number of ideas per author and period. The zero inflated model is a mixture model, where a logit and negative binomial process jointly explain the dependent variable. Columns (1) and (2) display the effect on the number of ideas per author and period. Columns (3) and (4) display the average marginal effects for the two processes separately. Columns (4) and (5) display the marginal effects for the logit and the negative binomial process, allowing the treatment effect to vary between *Prior ideators* and the rest. The unit of observation is the author-period, i.e., all ideas within a period of 13 months are summed up. Marginal effects of *Age* and *Tenure* are based on linear and quadratic terms. Standard errors are clustered at the client team level. ***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table 5: Treatment effect on different measures of idea quality

	(1) OLS	(2) Logit AME	(3) OLS	(4) Logit AME	(5) OLS
Dependent variable	Shared	Shared	Implemented	Implemented	Log(Net Value)
DID Treatment	0.209**	0.188**	0.153*	0.177**	0.308
	(0.087)	(0.079)	(0.084)	(0.087)	(0.390)
Number of Authors	0.082***	0.092***	0.060**	0.064***	0.185**
	(0.020)	(0.025)	(0.023)	(0.018)	(0.069)
Age	-0.036	-0.016***	-0.037	-0.012*	0.341*
	(0.025)	(0.005)	(0.025)	(0.006)	(0.182)
Age^2	0.000		0.000		-0.005*
	(0.000)		(0.000)		(0.003)
Tenure	0.044**	0.032***	0.014	0.008	0.129*
	(0.019)	(0.012)	(0.022)	(0.011)	(0.061)
$Tenure^2$	-0.002		-0.001		-0.005
	(0.002)		(0.002)		(0.006)
Male	0.038	0.037	0.114***	0.111***	-0.060
	(0.055)	(0.053)	(0.040)	(0.034)	(0.273)
Controls salary groups	yes	yes	yes	yes	yes
Controls project type	yes	yes	yes	yes	yes
Client FE	yes	yes	yes	yes	yes
Time FE	month	month	month	month	month
R^2	0.687		0.753		0.962
Log Pseudo likelihood		-546.45		-457.07	
Clusters	19	15	19	17	19
Observations	1779	1697	1779	1747	1912

Note: The table reports estimates of OLS and logistic regressions using as outcome variables the probability that an idea is shared with the customer (columns 1 and 2), the probability that an idea is accepted for implementation (columns 3 and 4), and the logarithm of the projected net value (profit from the idea) (column 5). The treatment effect is the difference-in-differences estimator. The unit of observation is the author-idea. Each observation is weighted by $1/(\textit{Number of Authors})$, where *Number of Authors* represents the number of employees who submit the idea together. Only ideas with finished review process (either accepted or rejected) are included the samples of columns (1) to (4). Marginal effects of *Age* and *Tenure* are based on linear and quadratic terms. Standard errors are clustered at the client team level. ***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table 6: Idea quantity and quality, treatment and post treatment effect

	(1) Zero Inflated NB Logit AME	(2) Zero Inflated NB Negative Binomial AME	(3) Quantity effect ZINB AME	(4) Quality Logit AME
Dependent variable	Pr(NonExcessZero)	NumIdeas w/o ExcessZeros	NumIdeas	Shared
DID Treatment	0.187***	-0.197**	-0.014	0.180***
	(0.062)	(0.094)	(0.040)	(0.066)
DID Post Treatment	0.255**	-0.160	0.026	0.019
	(0.106)	(0.130)	(0.056)	(0.202)
Age	-0.008***	0.006	-0.002	-0.012***
	(0.002)	(0.008)	(0.002)	(0.004)
Tenure	0.069***	0.009	0.028***	0.034***
	(0.009)	(0.015)	(0.003)	(0.010)
Male	0.053***	-0.090	-0.007	0.021
	(0.019)	(0.065)	(0.015)	(0.045)
Controls salary groups	yes	yes	yes	yes
Client FE	yes	yes	yes	yes
Time FE	period	period	period	month
Log Pseudo likelihood	-7767.68	-7767.68	-7767.68	-810.98
Clusters	19	19	19	16
Observations	25152	25152	25152	2310

Note: The table reports marginal effects for a zero inflated negative binomial model explaining the number of ideas per author and period, and of a Logit model explaining the probability of sharing an idea with the client. Marginal effects of *Age* and *Tenure* are based on linear and quadratic terms. Standard errors are clustered at the client team level. ***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table A: Description of variables

Variable Name	Description
Age	Employee age in years; measured at the end of the period (quantity regressions) or at time of submission (quality regressions)
DID Post Treatment	Difference-in-differences post-treatment effect estimate
DID Treatment	Difference-in-differences treatment effect estimate
DID Treatment Prior Ideator	Difference-in-differences treatment effect estimate for prior ideators
Finished	Dummy =1 if review is finished (idea either accepted or rejected)
Ideator	Dummy =1 if employee submitted at least one idea in respective period
Implemented	Dummy =1 if idea was accepted for implementation
Male	Dummy =1 if employee is male
Net Value	Profit in US dollar terms, projected value of idea minus projected cost
Number of Authors	Number of authors submitting the idea
Number of Ideas (NumIdeas)	Number of ideas (per employee) submitted in the respective period; the abbreviation is NumIdeas
Prior Ideator	Dummy =1 if employee submitted at least one idea in period 1 and is still employed in period 2
Salary Group	Set of dummy variables indicating an employee's salary level 0-8 (corresponds to hierarchical position in organization)
Shared	Dummy =1 if idea was shared with and communicated to client
Tenure	Tenure at the company in years; measured at the end of the respective period or at time of idea submission