# A Numbers Game: Quantification of Work, Accidental Gamification and Worker Productivity

**Aruna Ranganathan**
**Stanford GSB**

**Alan Benson**
**Carlson School of Management**

***August 31, 2019***

**Abstract**

Technological developments and the big-data revolution have facilitated fine-grained, high-frequency, low-cost measurement of individuals' work. Yet we understand little about the effects of this *quantification of work* on workers' behavior and performance. This paper investigates how and when quantification of work affects workers' productivity. We argue that quantification can affect worker productivity via *accidental gamification*, workers' inadvertent transformation of work into a game. We further argue that quantification is likely to improve productivity in a context of simple work, where quantified metrics adequately measure the work being performed; when work is complex, by contrast, quantification is likely to hurt productivity because quantified metrics cannot adequately measure the multifaceted work being performed, causing gamification to be demotivating. To substantiate our argument, we study the implementation of an RFID measurement technology that quantifies individual workers' output in real time at a garment plant in India. Qualitative evidence uncovers the accidental-gamification mechanism and three conditions enabling this mechanism; a natural experiment tests the effect of quantification of work on worker productivity. This paper contributes to the study of quantification, organizational practices and work design, and highlights important policy implications of increasing quantification of work.

New technologies and the big-data revolution have enabled the measurement of individual work performance (for example, see Anteby and Chan 2017, Christin 2018, Kim 2019). These technological developments facilitate high-frequency measurement of output that was previously hard to measure; they also reduce the cost of fine-grained measurement, prompting employers to measure all kinds of work (Bernstein 2017). Scholars of work have documented how jobs that once enjoyed considerable autonomy are increasingly shaped by, and even defined by, the intensive quantification of all aspects of the job. For instance, trucking companies are using GPS and onboard computers to track not merely routes and ride-completion times but also driving speeds and break patterns (Gray and Silbey 2014, Levy 2015). Similarly, IT programmers' code is increasingly being digitally measured using keystroke logging (Batt 2015). Such performance measurement can be accompanied by monitoring and surveillance (Lyon, Ball and Haggerty 2012, Marx 2016), but in this paper, we focus purely on the phenomenon of measuring work via the use of numbers and metrics, which we call the *quantification of work*.

Modern societies have seen a proliferation of measurements, rankings and benchmarks, usually adopted to promote efficiency, transparency and accountability (Porter 1995, Espeland and Stevens 1998, Berman and Hirschman 2018). The burgeoning sociological literature on quantification, defined as the "production and communication of numbers" (Espeland and Stevens 2008: 402), observes that numbers wield broad influence as sources of "truth" and "rationality" (Mazmanian and Beckman 2018). Scholars of quantification have demonstrated that quantified metrics are transforming multiple fields, from education (Sauder and Espeland 2009, Sharkey and Bromley 2015) to non-profits (Hwang and Powell 2009), the credit market (Kiviat 2017, Fourcade and Healy 2017), criminal justice (Brayne 2017), and everyday life (Lupton 2016).

Scholars of quantification point out that numbers have the potential to alter the trajectories of individuals, organizations and entire fields of endeavor (Espeland and Sauder 2016). Three main mechanisms have been posited to explain why individuals, in particular, might change their behavior in response to quantification. First, quantification might incentivize strategic behavior that maximizes measured aspects of performance at the expense of unmeasured aspects (Sauder and Espeland 2009). Second, quantified data are amenable to computational analysis, or what has come to be called big-data analytics, that could enable organizations to direct employees through real-time decisions (Brayne 2017). Finally, quantification could give rise to new regimes of control that lead individuals to modify their behavior in response to disciplinary action or the threat of discipline (Levy 2015).

Irrespective of the mechanism at play, the literature emphasizes that quantification tends to evoke standard reactive practices from individuals because quantification is essentially a vehicle of *commensuration*—that is, of evaluating different persons, processes, locations and artifacts with a common metric (Espeland and Sauder 2007). By deliberately stripping away context and social relationships to render different events similar, new regimes of measurement provoke individual responses that tend to converge and to take similar forms across institutions using the same set of metrics (Espeland 1993, Kiviat 2017, Berman and Hirschman 2018). Thus most sociological studies tend to emphasize the connection between quantification and standardized reactions (see Christin 2018 for an exception).

In this paper, we study a case of workplace quantification and investigate how it affects productivity within a homogenous group of workers. Though none of the three recognized mechanisms—incentives, analytics and discipline—is at play in this case, quantification improved some workers' productivity but impaired that of others. This pattern poses a challenge to the existing literature on quantification, which has focused on the three aforementioned channels and has predicted convergence in individual responses. Motivated by this puzzle, our key research question is: how and when does quantification of work affect worker productivity? Our question also responds to calls from researchers who note that "investigations of measurement can be advanced by further specifying the mechanisms and effects of reactivity" (Espeland and Sauder 2007: 34).

We argue that quantification of work can affect worker performance via "accidental gamification," a mechanism that we define as workers' inadvertent transformation of work into a game: When numbers introduced by quantification provide (a) clarity of objective, (b) competition with oneself, and (c) interactive feedback, workers may not even realize that they are slipping into an addictive game of "binge working." We further argue that quantification is likely to improve worker productivity in the context of simple work, where gamification is motivating because quantified metrics adequately measure the work being performed. When work is complex, by contrast, quantification degrades worker productivity because quantified metrics cannot adequately measure the multiple dimensions of the work being performed, causing gamification to be demotivating.

To develop this argument, we study the pilot implementation of a radio-frequency identification (RFID) technology that transformed previously unquantifiable individual worker output into real time quantified metrics at a large garment-manufacturing plant in India. Specifically, we investigate how this quantification of work affected the productivity of a similar group of assembly-line workers quasi-randomly assigned to tasks of varying complexity. The RFID technology was implemented in late 2012 in three of twelve production lines; in each production line, workers perform a specific task or operation. The treated lines were chosen for a reason unrelated to productivity. Work-in-process garments were tagged with RFID tags; workers were instructed to scan a garment's tag on the RFID scanner newly installed at their workstations prior to working on it, thus enabling the quantification of individual productivity in real time, visible to management and to the workers. As a pilot, the quantification of work was not tied, nor perceived by workers as tied, to rewards or punishments, nor were any analytics performed on the quantified performance data.

This paper is organized around the full-cycle research model (Fine and Elsbach 2000), which mirrors the research process that we followed. First, we conducted one year of ethnographic fieldwork that provides evidence for the mechanism of accidental gamification as well as delineates two hypotheses specifying how quantification of work affects productivity for simple and complex work respectively. Next, to test the two hypotheses around productivity that emerged from fieldwork, we collected six years of daily line-level productivity data before and after introduction of RFID measurement to jacket and pant lines, as well as daily operation-level productivity data for a sample of days before and after introduction of RFID measurement to pant lines. Access to such pre- and post-intervention data on treated and nontreated groups is an asset. We further exploited variation in the complexity of pant and jacket lines—production of pants is rated as less complex than that of jackets—and in the complexity of different operations within pant lines (to address possible objections that pant and jacket lines could differ on dimensions other than complexity). A key advantage of our setting is the opportunity to exploit minute differences in work complexity rather than comparing vastly different kinds of work.

Our study makes three contributions to the literature on quantification. First, we show that quantification, in itself, and without relying on any of the mechanisms emphasized in the prior literature (incentives, analytics or discipline), can affect worker behavior and productivity.

Second, we demonstrate that quantification of work can affect worker behavior via the novel mechanism of accidental gamification for which we uncover three enabling conditions. And third, we show that quantification of work does not always evoke convergent reactions but can produce divergent effects by factors such as work complexity. Our study also contributes to the literature on gamification in organizations by finding that games can sometimes increase and at other times, decrease worker productivity even when they are not strategically introduced by organizations. We further contribute to the literature on work design by demonstrating that work complexity interacts with new forms of technologically mediated quantification to affect workers' motivation and productivity.

## QUANTIFICATION OF WORK: MECHANISMS AND EFFECTS

Sociologists have drawn attention to the rise of quantification and the significance of new regimes of measurement (Espeland and Sauder 2008, Espeland and Stevens 2008, Berman and Hirschman 2018), arguing that numbers offer a shared language of quantity while erasing the local and the particular (Porter 1995). They posit that, once numbers are gathered, they often acquire an aura of objectivity, and travel easily across time and space (Christin 2018).

Quantification has been studied in many contexts. Audits, standardized tests, scorecards, and a wide array of ranking regimes all represent cases of quantification designed to evaluate individuals and organizations and to make them more accountable (Sauder 2008, Sharkey and Bromley 2015). College rankings, in particular, have attracted significant academic attention (for a review, see Espeland and Sauder 2016). Scholars have also studied "results-based accountability" and other quantitative metrics that are increasingly being used in the non-profit sector (Hwang and Powell 2006, Keevers et al 2012). In the personal domain, scholars have studied the plethora of tools to quantify our steps, meals, heartbeats and other aspects of our personal lives (Lupton 2016, Neff and Nafus 2016).

The workplace domain too is experiencing a trend toward quantification. Management scholars have noted that tools for budgeting, forecasting and strategic planning can function as rituals of quantification in organizations (Miller and Power 2013, Mazmanian and Beckman 2018); other scholars have investigated the use of quantified data to make decisions (Karunakaran 2017). Less attention has been paid, however, to quantification of individual workers' output or to how workers react to quantification of their work. (For a notable exception, see Christin's (2018) investigation of quantification in the field of journalism, in the form of "clicks" on online articles.)

Organizations have long measured worker output to implement piece-rate incentives (Lazear 2000), monitor workers (Hubbard 2000, Pierce, Snow and McAfee 2015, Staats et al 2017) and to provide performance feedback (Greve 2003), and recently, technological developments have made it even easier to engage in high-frequency, fine-grained measurement of individual workers' performance (Bernstein 2017). However, existing scholarship has focused less on the pure effects of quantification or measurement of work, stripped away from the add-on treatments of monitoring or rewards. Because we still lack insight into "how such [quantification] processes operate in practice" (Mazmanian and Beckman 2018: 377), there have been calls to study, for example, how "professional norms, work practices, and organizational dynamics shape the impact of digital technologies" (Christin 2018: 1411). Responding to these calls, this paper asks: how and when does quantification of work affect workers' productivity?

**How Quantification Affects Worker Performance**

Quantification scholars have theorized that people change their behavior in response to being evaluated, observed and measured, and have explored why such "reactivity" occurs (Espeland and Sauder 2007). The literature on quantification has specified three mechanisms through which quantification could influence individual behavior.

First, quantification might change incentives or create new incentives that elicit strategic behavior (Sauder and Espeland 2009). Such incentives could be financial, reputational or even identity-based (Anteby 2008). For example, Espeland and Sauder (2007) documented that, once quantified college rankings were created, some deans were explicitly tasked with achieving higher rankings. In an effort to improve their schools' rankings, these deans changed how revenue was allocated, how scholarships were awarded and how graduates found jobs (Espeland and Sauder 2007). Some even made cynical efforts to manipulate ranking data, such as by misrepresenting graduates' employment rates rather than actually placing more graduates with employers (Sauder and Espeland 2009). While not theorizing about quantification per se, Burawoy (1979) similarly described a manufacturing firm that quantified workers' output to pay them at a piece rate, noting that workers responded by producing at a pace defined as "making out" to maximize their wages while gaining some control over the terms of their work.

Quantification could also change behavior via big-data analytics performed on newly available data (Pentland 2014). Various organizations have hired data scientists to perform computational analyses of digital datasets consisting of high-frequency quantitative observations (Karunakaran 2017). Big data is fast becoming a tool not just to analyze patterns but also to predict the likelihood of an event (George, Haas and Pentland 2014, Brayne 2017). This tool creates new knowledge that enables organizations to direct employees based on real-time decisions about business operations (McAfee et al 2012); individuals could change their behavior in response to new patterns detected, predictions generated or knowledge created.

Finally, discipline could drive reactive practices in the wake of quantification (Christin 2018). Quantification has long been seen as a tool for control (Porter 1995, Foucault 1997); it expands the comparability of social phenomena in ways that permit strict discipline (Espeland and Stevens 2008). Studies have shown that performance measurements can be mobilized as panoptic technologies of surveillance: individuals may not know who is tracking their performance at a given moment, or why, but try to perform well nonetheless (Sewell 1998, Levy 2015, Anteby and Chan 2017). Sometimes administrators monitor quantified performance and discipline poor performers, prompting such individuals to modify their behavior (Christin 2018); explicit disciplinary pressures can also devolve into self-management (Covaleski et al 1998).

All three mechanisms—incentives, analytics and discipline—help to explain the effects of quantification on individual behavior, but none of them is at play in our setting. At the garment plant we study, quantification was not paired with any real or perceived financial or reputation incentives. Nor did management hire data scientists or task anyone with performing big-data analytics. Furthermore, front-line supervisors did not look at the data or use it to discipline workers. Even so, quantification seems to have affected worker productivity. In order to understand this phenomenon, we will turn to the management literature on gamification in organizations.

**Gamification in the Workplace.** An emerging literature defines gamification as "firms consciously apply[ing] games and game dynamics to an extremely wide range of uses, from innovation to recruiting" (Mollick and Werbach 2015: 439). This deliberate adoption of gamification as a strategic management tool at enterprises outside of the gaming sector (Edery and Mollick 2009) has coincided with the rise of a "gamer generation" that grew up with videogames (Beck and Wade 2004).

Examples of enterprise gamification to increase worker productivity abound. In 2015 Microsoft introduced a game called GamEffective at its customer-support call center to improve agents' engagement, satisfaction and retention. The game included level advancements, badges and scoring systems (Frost and Sullivan 2016). To make everyday business tasks more engaging, a growing number of firms, including IBM and Deloitte, are incorporating elements of videogames, such as awarding points for meeting deadlines, into the workplace (Silverman 2011). Firms like Badgeville have come into being to offer gamification services to established organizations (Doughtery and Hardy 2015).

Empirically, recent studies seem to suggest that gamification improves worker productivity. Mollick and Werbach (2015) studied an organization that implemented gamification to improve the performance of its 20,000 call agents and observed that the system reduced agents' call times and improved sales. Similarly, using a field experiment at a startup, Mollick and Rothbard (2014) found that "scoreboard" games, when consented to by salespeople, increased their sales performance. Other studies have shown that game-oriented practices can increase worker productivity (DellaVigna 2009, Blanes i Vidal and Nossol 2011). Scholars posit that gamification makes work more engaging and taps into the human desire for fun (Deterding et al 2011, Zichermann and Cunningham 2011). As game developers have found, a fun process coupled with a system of incentives for a job well done can become downright addictive (Mangalindan 2010).

To date, the literature on gamification has not considered employees' inadvertent gamification of their own work, or its effect on productivity. Early sociological research documents some examples of employee-initiated games such as shopfloor workers playing poker at work or stealing and hiding each other's bananas (De Man 1928, Roethlisberger and Dickson 1943, Roy 1952). Roy (1959:158) explains this tendency to play games as a means to relieve the "beast of monotony" that sets in when work is so exceedingly tedious that it would make any worker "go nuts." Such games are further encouraged by managers because they create "consent" to the work (Burawoy 1979). However, these games are consciously generated by workers to relieve monotony, are not necessarily related to quantification and it is unclear whether they impact productivity in a meaningful way (Roy 1959, Burawoy 1979).

The pattern of "accidental gamification" that we observed during fieldwork is quite distinct. As we will show using our qualitative data, quantification of work offers a defined numerical outcome, an avenue for competing with oneself, and an interactive feedback system, and can thus facilitate what we call "accidental gamification." This mechanism might help explain how quantification affects worker behavior in the absence of incentives, analytics or discipline.

**Variation in the Effects of Quantification**

This paper also investigates quantification's actual effects on productivity. The literature tends to emphasize that quantification evokes standard reactions from individuals because it inherently entails commensuration: that is, it transforms qualities into quantities that share a metric (Espeland and Stevens 1998). Commensuration unites objects and people by incorporating them in a shared cognitive system that expresses difference and similarity in terms of magnitude (Espeland and Sauder 2007). Quantification provokes similar reactions from individuals facing the same metrics because commensuration erases alternative definitions of "what counts" and imposes a standardized, universal definition of "high quality" (Fourcade 2011, Lamont 2012).

Espeland (1993) describes how the quantification of credit scores has commensurated different lived experiences of debt and default, erasing distinctions between delinquencies that professionals considered legitimate and those that they viewed as unacceptable; as a result of the "stripping away of context," different professionals evaluating the same set of credit reports react very similarly and arrive at very similar conclusions (Kiviat 2017). The only circumstance in which quantification has been shown to produce divergent responses is in different national contexts where interpretations of metrics differ (Christin 2018).

This paper, however, presents a case of workers in a single cultural context—in fact, a single organization—reacting differently to the quantification of their work. To explain this, we argue that a key question is whether the quantified metrics are a reasonable reflection of the underlying work being measured. When the validity and reasonableness of metrics are accepted by those whose accomplishments or worth they purport to measure, such measures can positively influence the activities being measured; otherwise, measures can render some aspects of life invisible or irrelevant. This paper argues that, in order to understand the effects of quantification of work, we need to pay attention to whether there is a close link between the newly quantified metric and workers' pre-existing notions of performance. We will turn next to the job design literature in industrial and organizational psychology for insight into *when* there is likely to be a close link between a newly quantified metric and prevailing notions of performance.

**Job Design and Work Complexity.** Scholars have long asserted that complexity is an important characteristic of work (Edwards, Scully, & Brtek 1999, Parker 2014, Parker, Van den Broeck and Holman 2017, Hackman and Oldham 1976). Defined as the extent to which work is multifaceted and difficult to perform (Humphrey, Nahrgang and Morgeson 2007), complexity is understood as the degree to which work in its very substance requires thought and imposes cognitive demands on the individual who performs it (Kohn and Schooler 1978), by contrast to job simplicity (Campion 1988, Edwards, Scully, & Brtek 2000).

Job-design scholars have theorized, and empirically shown, that work complexity affects motivation and productivity. Because work that involves complex tasks is mentally demanding and challenging, it is sometimes posited to have positive motivational outcomes (Morgeson and Humphrey 2006, Humphrey, Nahrgang and Morgeson 2007). For example, Frese, Garst and Fay (2007) showed that job complexity predicts a control orientation (a motivational state that includes self-efficacy). Other research has suggested, however, that job complexity is also likely to erode efficiency and promote perceived work overload. Xie and Johns (1995) found that both high and low job complexity were related to high levels of exhaustion; moderate levels of complexity were not. Edwards, Scully, & Brtek (2000) also found job simplicity to be positively related to self-reported efficiency. In other words, high-complexity work both engages and overwhelms job holders (Humphrey, Nahrgang and Morgeson 2007).

While we have some understanding of the direct effects of work complexity on worker behavior, scholars who theorize about work complexity have begun to point out important changes in the nature of work that have received little attention. For example, new initiatives enabled by information and communications technologies allow more systematic collection of outcome data (Van der Spiegel 1995, Morgeson & Campion, 2003, Humphery, Nahrgang and Morgeson 2007). One such organizational initiative is the quantification of work (Vough and Parker 2008). Scholars are investigating the intermediate role of work design in efforts to understand the effects of phenomena like lean production (Parker 2003) and teleworking (Feldman & Gainey 1997), but have noted that the "interactions between performance [tracking] and work characteristics have received scant research attention" (Parker, Wall and Cordery 2001: 422).

Our paper addresses this gap. We investigate how workers performing more and less complex jobs react to quantification of their work, and how their productivity changes as a result. We argue that it is easier to devise acceptable quantified metrics for simple work than for complex work, which is more multidimensional. We posit that workers who perform simple work will therefore react positively to quantification, resulting in an improvement in their productivity, and that those engaged in complex work will react negatively to quantification, reducing their productivity. The section that follows describe our setting.

## SETTING: A GARMENT MANUFACTURING PLANT IN INDIA

Our setting is a plant in India that manufactures men's suits. The plant's products pass through a linear sequence of non-divisible operations; the plant features 12 production lines and 147 operations within those lines, and employed 2,212 line workers during the study period, 2009–2014. Each line produces one type of product and consists of a set of operations: the 9 pant lines each consist of 51 pant operations; the 3 jacket lines each consist of 96 jacket operations. Two advantages of our setting are that the lines' configurations did not change during the study period and the line and operational assignments were based on operational need, rather than person characteristics, so lines and operations are similar in terms of demographics and measures of skills. Workers perform only one operation, but the complexity of those operations varies widely.

In 2012, plant management introduced digital performance-measurement technology to a few treatment lines on a trial basis, to measure workers' production and to help them meet deadlines. The CEO had seen this technology in use at plants in China and was keen to implement it. Workers were told nothing about the purpose of the intervention; the scanners were installed overnight, and the next day workers were taught to use them. Unfinished items bore RFID tags; scanners were installed at workstations, and workers were instructed to scan a garment's tag before working on it. Thus its progress could be tracked down the production line. Scanners reported the number of units produced and individual efficiency as a percentage of set targets in real time. Figure 1 shows an RFID terminal in the plant.

[Figure 1]

Thus the intervention in this case should be thought of as a quantification-of-work intervention that digitally measured individual workers' productivity in real time, in a way that was visible to management and to workers. Previously, only line-level productivity could be measured at the end of a day, by counting the number of pieces produced, but information on how an individual worker was performing was hard to measure. The intervention enabled the quantification of this individual worker productivity at every moment of the day.

Despite the possibilities enabled by quantification, surprisingly none of the quantified data was used in any systematic way to design rewards, perform data analytics or discipline workers. This is explained by the fact that the intervention we study was a limited pilot rollout of the RFID technology that was eventually never expanded to the rest of the factory given prohibitive cost. Therefore pay remained unchanged and workers continued to be paid a fixed daily wage. Interviews with workers suggest that they did not perceive there to be any long-term incentives either. Additionally, the plant did not hire data scientists or task anyone with performing big-data analytics on the quantified performance data.

Further, immediate supervisors did not use the quantified data to discipline workers. Supervisors divided their time among several lines, and the new data was only available for treated lines. Supervisors also lacked easy access to the data; the RFID data were only accessible via a computer on a different floor. Inertia, too, discouraged abandoning or supplementing the traditional system of face-to-face supervision and discipline.

In sum, any effects that quantification of work had on worker productivity in this setting would have entailed a mechanism other than incentives, analytics or discipline.

## FULL-CYCLE RESEARCH METHODS

To investigate how and when quantification of work affects worker productivity, we adopted a full-cycle research design. This approach combines inductive and deductive methodologies, in a cyclical manner, in a single research program (Fine and Elsbach 2000, Ranganathan 2018). The logic is that initial qualitative data can richly describe real-world issues that are worth studying, and can generate theory and hypotheses that grow directly out of the immediate experiences of informants; quantitative data can then identify generalizable causal relationships. Our ethnographical fieldwork and interviews at the garment plant generated our theoretical mechanism of accidental gamification and two hypotheses about how work complexity moderates the effects of quantification. Then, we tested our two hypotheses using unique longitudinal productivity data.

### Qualitative Methods

We performed ethnographic fieldwork from June 2014 to June 2015, and conducted 41 interviews between January and June 2015. This qualitative data, collected by one of the authors and a research assistant, produced over 200 single-spaced pages of fieldnotes and interview transcripts. The fieldwork and interviews were conducted post-intervention, after the quantification technology was installed on the treatment lines. It would have been ideal to observe the intervention as it unfolded, but collecting qualitative data 1.5 years after implementation of the RFID measurement system allowed us to capture workers' "equilibrium" response to quantification of their work, rather than fleeting, immediate short-term reactions.

Ethnographic observation consisted of documenting daily activities in the pants and jackets sections, and observing how workers were supervised on both treated and nontreated lines. Each day of observation consisted of choosing one line, and one sub-section of it in particular, to focus on. This approach enabled us to hone in on a small group of about ten workers for an entire day; we became thoroughly familiar with those workers' operations and daily routine, their interactions with their supervisor, the pressures of meeting the day's work demands and their means of coping with those pressures. When observing workers on treated lines, we paid attention to how they interacted with the RFID technology, and to the aspects of the technology that they appreciated and were frustrated by. While observing, we chatted with workers in Kannada and Tamil (the languages spoken on the shopfloor). Workers were eager for their stories to be told, and seemed to open up to us because we asked about their work and documented their successes as well as their frustrations. We always carried conspicuous notebooks and let workers see us jotting notes. At the end of each day, we made sure we had documented all salient observations.

We structured our time to be in the field approximately three of the six workdays[1] per week; we spent the rest of the time typing up fieldnotes, writing memos and interpreting the emerging data. Time spent away from the shopfloor enabled us to identify puzzling observations that would then guide what to focus on in the following week's fieldwork. By the end of the fieldwork, we had observed all twelve lines and all workers engaged in tailoring operations.

[1] Most organizations in India subscribe to a six-day workweek; Sunday is the weekly holiday.

We also conducted forty semi-structured interviews: 26 with workers (W1–W26), 4 with front-line supervisors (S1–S4), 4 with industrial engineers[2] (IE1–IE4) and 6 with plant executives (H1–H6). See Appendix A for more detail about the interview sample. This sample captured diverse perspectives and enabled us to triangulate across varied responses. The aim of the interviews was to understand how various stakeholders described, understood and responded to the RFID performance-measurement technology. The interviews were conducted in Kannada, Tamil or English, and lasted, on average, one hour; each was digitally recorded. After every interview, we recorded our impressions of the interviewee, and of the physical surroundings and conditions under which the interview was conducted. We analyzed this qualitative data inductively to develop our hypotheses (Glaser and Strauss 1967, Strauss and Corbin 1990). Analysis consisted of multiple readings of field and interview notes, composition of analytical memos and tracking of patterned activities and issues over time.

## QUALITATIVE DATA AND HYPOTHESIS DEVELOPMENT: ACCIDENTAL GAMIFICATION AND WORK COMPLEXITY

### RFID Measurement Technology: A Case of Quantification of Work

"Why implement RFID?" (H5) the head of the plant asked rhetorically, repeating the question we had posed to him. "It is important to understand how many pieces each worker is producing in the 480 minutes of each day.… Each minute lost is in turn equal to a piece not produced." Our interviews with senior management suggested that the tracking system had been introduced to quantify the work performed by individual workers. RFID "gives you worker-specific information and helps in tracking pieces" (H1), the head of HR added.

[2] Industrial engineers are commonly employed at manufacturing facilities and tasked with improving efficiency in production processes.

A key feature of the RFID system was that it enabled "quantitative data collection" (H2). As scholars have pointed out, numbers enjoy considerable sway as embodiments of trust, objectivity and rationality; management at this plant too put faith in the power of numbers. Executives asserted that quantified data "helped in approaching matters in a more *systematic* way" (H3) and allowed for "*accurate* numbers" (H2) [emphasis added]. Comparing RFID-enabled data to what had existed earlier, the head of production said:

> Earlier we had no idea. Our best estimates [for what] individual workers were producing were round figures; numbers like 63 or 72 were never in the data.… [but] the data from RFID helped know the actual output of a worker, in contrast to what was earlier assumed about them. Workers are also usually in the habit of assuming they produce the same numbers every hour; but with RFID this [too] was overcome (H4).

Despite introducing an RFID system that enabled the collection of precise quantitative data, senior management chose not to link the RFID system to financial incentives. According to the head of production, "Our setup here is a fixed-salary environment, where there is no piece-rate system" (H4). An industrial engineer concurred: "If the plant had plans to roll out an incentive system after the pilot, the RFID data will be of great use … [as it] gives the accurate numbers per worker in the line … [but] there are no plans in the works" (IE3). Senior management did, however, expect front-line supervisors to monitor the data on workers' productivity and to discipline workers who performed poorly. The production head for pants explained: "Ideally, how the RFID should be used is the worker productivity data should be monitored, analyzed, [and] problem workers should be identified and reprimanded" (H4). In reality, the RFID system was not associated with supervisory discipline. In response to our question "Do supervisors look at the RFID data?," an industrial engineer said: "They should be using the data but they don't" (IE1). In fact, when we went to collect individual productivity data for our research, we found that it had never been downloaded to the server![3] Given that this RFID program was a pilot, senior executives were not motivated to respond to this oversight by reprimanding the supervisors, or to hire data scientists to perform big-data analytics on the newly available quantitative performance data.

Workers also knew that the RFID system was not associated with incentives, analytics or discipline. Interviews revealed that they had no illusions that the RFID data was being monitored systematically or would be used to punish or reward them. One worker said, “Unless someone actually monitors and acts on my data, I don’t have an incentive to be faster. Currently, nobody seems to be actually doing this. So the workers don’t bother about the system” (W19). Another worker said, “I do not remember supervisors ever coming, checking the machine and reacting to it” (W2). An HR manager corroborated these observations: “If there were incentives linked to [the RFID], and workers knew about that, maybe it [RFID] would be seen as a positive thing because their payment would be attached to it. But workers know there are no incentives” (H1).

Despite being unassociated with incentives, analytics or discipline, workers’ productivity was affected by RFID-enabled quantification. “The system seemed to be giving production improvements on its own” (H2), the head of pants production observed. The head of production elaborated: “An ECG, by itself, will not cure a disease, but merely acts as a tool which helps in diagnosis; similarly, however good a car may be, a good driver is critical to use it…. I thought the RFID would be like an ECG or a car, but the RFID seems to be different. Even without the technical team using it, simply installing the RFID machines seemed to improve production” (H4). This puzzling observation sets the stage for our analysis that uncovers how quantification of work might affect worker productivity.

[3] We were the first to ask for the data. It took over a month for a data clerk at the plant to coordinate with the manufacturer of the RFID devices and obtain the data for us.

**Accidental Gamification of the RFID Measurement Technology**

We found that the very act of quantifying workers' output lent itself to gamification. The new availability of RFID numbers seemed to promote inadvertently treating work like a game, a phenomenon we call *accidental gamification*. Workers could see their live production numbers and "current efficiency" on their RFID devices; they reported that "earlier it was hard to track production but the machine had made this easier" (W14). They described "slipping into a game without even realizing it" (W1) and then getting "hooked to it" (W2), and "smash[ing] their way" (W12) through tasks; we noticed that they often talked about their production numbers with "delight on their faces" (fieldnotes). By gamifying work, they made time fly and "work became fun" (W2).

Our fieldwork identified three features of the RFID system—(a) clarity of objective, (b) competition with oneself, and (c) interactive feedback—that prompted workers to gamify their work. First, workers seemed to define, for themselves, a clear objective that signified a win. The objective of their game was "getting to 100%" (W10) in daily efficiency. One worker said, "I am waiting for a day when my current efficiency becomes 100%!" (W13). At the time this worker's efficiency was about 60%, in line with the plant average. Workers had been aware earlier that higher efficiency numbers were better, but hadn't known where they stood, and 100 had not been an objective they were actively striving toward. Ironically, 100% efficiency was an unachievable target. "While we set a target of 100%, our expectation is only 80%" (IE1), an industrial engineer told us. Workers were unaware of this, resulting in a game that was quite hard to win at.

Second, workers seemed to compete with themselves.[4] The RFID devices seemed to breed comparisons between a worker's past and current performance. They also shaped goals for

[4] Competing with others was less likely: productivity numbers were not made public and the numbers displayed on a particular worker's screen were hard to see from a distance.

future performance. One worker said, “Today if I do 80%, then tomorrow I try for 81%. . . . In this way, each day is a competition!” (W8) During fieldwork, a worker told us, according to our fieldnotes, that “she usually manages to get her production numbers to around 230.” Later the same day, we observed that “her machine was displaying only 179 [and that] she worked with utmost focus trying to get her numbers up until the plant bell rang indicating that it was 5:30 pm.” Another worker told us “that the last hour was her time to push the maximum number of pieces” and “outdo herself” and that she “relished seeing her RFID machine at this time” (W10).

Third, workers seemed to interact continuously with their RFID devices. Workers frequently scanned RFID tags that changed the numbers on their screens. One worker said that looking at “the RFID data real-time is like following cricket scores live.” He added that the “machine gives every possible detail about a worker,… like their *jathakam* (horoscope)” (W22). Another worker reported that she glanced at her efficiency number each time she scanned a tag and that “she knew that she has to start working faster when the efficiency number dropped” (W8). The plant manager speculated that “most of the workers were not really educated, and so they found the interactive devices especially fascinating” (H5). Many workers said they were “interested in gadgets” (W14) and liked “the buttons on the RFID device” (W5); we observed that some workers “had [even] taken the initiative to stitch beautiful covers for their RFID scanning panels on their own time.”

We thus posit that quantification affected worker productivity via the mechanism of accidental gamification. The RFID technology offered a clear objective, competition with oneself and interactive feedback, prompting workers to gamify their work of their own accord, without interference from management.[5] Table 1 offers additional evidence of accidental gamification.

[Table 1]

However, we also found heterogeneity in workers' opinions on whether or not accidental gamification was motivating. Next we will explore the conditions under which accidental gamification was seen as motivating and resulted in improved productivity.

**Heterogeneity in the Effects of Quantification by Work Complexity**

Observation revealed significant heterogeneity in the complexity of the work being performed on the shop floor, though all workers were engaged in garment production. Complexity was understood as the time required to perform a specific task, once tasks had been carefully engineered so that they could not be broken down further. Thus, simple operations took less time to execute than complex operations, reflecting the fact that complex tasks were more multifaceted and difficult than simple ones. Our interviews revealed that work complexity was an important factor in how workers understood their work and in their interpretations of and reactions to the digital-measurement technology.

**Simple Work.** Workers who performed simple work understood their work as "routine" (W1); their objective was to get it done fast. One said, "I do a straight stitch along the seam.... There's no thinking. I just need to keep the production going" (W13). Another said, "All I have to do is use my arms to align my pieces. The machine does the rest…. I could do this in my

[5] Accidental gamificaition did not seem to occur in the nontreated lines. Our fieldnotes document that "some workers in the nontreated lines tr[ied] to keep track of their output by scratching or writing numbers on their work desks and these workers' desks [were] very messy and crowded with all the count markings." In response to questions, these workers lamented that their makeshift system was mostly unsuccessful because they were prone to "counting mistakes" (W2) and often "goofed up because they were illiterate" (W17).

sleep; it's that easy" (W5). Workers who did various simple operations called their jobs "boring" (W13), "repetitive" (W8) and "monotonous" (W7).

Because such work was simple, the quantity metric captured by the RFID devices on the treatment lines—the number of pieces produced—was seen as an appropriate and useful measure of performance. One supervisor said, "Pants are produced in huge quantities.…Higher the quantity, better the pants production…. So with a technology like this, one could see a payback" (S3). The head of production concurred: "Number of pieces produced is the right indicator of performance in the case of products where there is mass production, like pants" (H4). A worker who performed a simple coin-pocket operation said, "With RFID, I am now able to see how I am working; I am able to see my quantity and percentage efficiency. I have a target of 35 pieces per hour for my current operation, and I know that right now I am able to deliver 25–30 pieces" (W8). Because such workers saw quantity as a meaningful measure of performance, they viewed accidental gamification as motivating. The head of production explained:

> Earlier, workers doing simple tasks felt like they were traveling on a long, straight toll road with no distance markers, no sense of how far they've come. Now those who have RFID have distance markers, and they look forward to seeing the distance they've covered. They know how far they need to go, and they, without realizing, try to cover the distance faster than the previous time they travelled on this road" (H4).

The sentiment that the technology had enabled a game that made simple work fun was articulated by a worker: "I have been in the plant for many years and have been doing the same operation from Day 1. I do not like the operation very much, but I like the RFID machine and I like to do the scanning…. It's like a calculator, and my goal is to make sure that efficiency does not go down in a day,… because, once efficiency goes down, I've lost—game over—and it is very difficult to bring it up" (W5).

Observation further suggested that the productivity of workers who performed simple work had been positively affected by the RFID measurement technology. The worker who described the machine as a calculator said that “she thinks that this has helped her work faster and deliver more production” (W5). A worker engaged in basic assembly loading said, “The machine has helped me realize my potential. . . . I am delivering better production numbers than I ever thought I could do!” (W4) A senior worker explained, “The production definitely improved.… It has now become part of the workers’ work” (W15). A supervisor on the treated pants line noted that “After RFID, the line is running very strong with workers displaying higher degree of self-interest” (S3). Based on our qualitative data, we hypothesize:

**Hypothesis 1**. *In a context of simple work, the effect of quantification of work on worker productivity will be positive.*

By contrast, observation indicated that workers engaged in complex work responded very differently to quantification.

**Complex Work.** Workers who performed complex work saw their work as “craft” (W22) and focused on getting it “done right” (W26), rather than just quickly. Such workers described their jobs as “fulfilling” (W24) and “challenging” (W14). One worker asserted that “jackets were premium products and almost every piece felt customized” (W15). Another elaborated: “A jacket is seen as a product that would add esteem to a person wearing it,…and whether or not that happens is in my hands” (W114). A worker who performed a complex sleeve/headroll-attach operation said, “I stitch pieces of fabric varying in size onto a curved piece of felt material which supports the sleeve across the shoulder—it’s an accomplishment every single time!” (W24) A worker who performed a complex armholes operation even reported that "she enjoys her operation so much that she did not take much leave even for her own wedding" (W26); she commented that "her relationship with her work was that deep" (W26).

When RFID was implemented on the treated complex lines, the quantity metric it captured was seen by workers as a partial and thus imperfect measure of performance. The head of production explained that "in jackets, the drive is towards 'premiumness' and quality, and not so much numbers produced" (H4). The head of jacket production concurred: "Though there is some drive toward production, the drive for quality is stronger" (H3). A worker elaborated: "Side seam is a complex operation. I think it is complex because it is very difficult to do. If I make even the smallest error, a wrinkle will form and the piece will come back for alterations. So number of pieces is not the only thing [that matters]; whether there are wrinkles or not is just as important" (W22).

In this context, quantification of work output induced accidental gamification that sapped motivation. One worker said, "Before the machine was introduced, I was able to allocate my time on my own, but now the countdown has added additional work pressure . . . and made it not fun" (W20). Asked to elaborate on the countdown, she said, " It is hard not to get sucked into the RFID countdown game, but you know how some games are fun and others aren't? This one isn't, for me" (W20). Workers reported feeling demotivated and devalued because RFID had the effect of "reducing them to a number" (W22). They took pride in their craft and wanted to be appreciated for delivering high-quality products. One worker who sewed intricate internal pockets said, "Workers like me deliver because of our deep familiarity and interest in our operation,… not because of games you put on my desk" (W25). Thus workers engaged in complex operations also inadvertently turned their work into a game, but found it neither enjoyable nor motivating.

Furthermore, this reaction to quantification among those doing complex work seemed to result in lower productivity. A worker who performed the under-collar zigzag stitch said, "I do not like having the watchman machine. . . . I used to really like my operation. Now I'm not so sure. . . . [I] think that it has not positively impacted my production levels" (W27). A worker who performed a complex lapel-seam operation said, "My work is critical and difficult to do. . . . [Being] observ[ed] while work is being done would make even a normally fast worker slow down" (W23). A worker who performed a sleeve-lining operation said that, though "the target for the operation is 70, I am able to deliver only about 55–60 pieces now, after the RFID" (W20). Another worker said that "in general, she does not seem to like having the machine. She thinks the scanning has actually brought down her numbers" (fieldnotes). Thus we hypothesize:

**Hypothesis 2.** *In a context of complex work, the effect of quantification of work on worker productivity will be negative.*

Table 2 presents additional evidence that quantification differentially affects workers engaged in simple and complex work.

[Table 2]

Note that even though the reactions of workers performing simple and complex work were quite different, we observed that the workers themselves were quite homogenous: most came from villages surrounding the factory, were economically disadvantaged (prompting them to seek this job) and had not previously been formally employed. Further, there appeared to be no systematic assignment of workers with certain characteristics to certain jobs. On the contrary, jobs were assigned to workers on a quasi-random basis based on the location of vacancies irrespective of whether the work is simple or complex. Specifically, we observed that a handful of workers "walk in" to this plant on a daily basis seeking jobs as sewing operators. The factory did not use a recruitment test, behavioral interview or other screening procedure. Most workers who walked in were hired the same day—the only criteria for rejection were color-blindness and inability to read numbers—and assigned to an open job (a particular operation on a particular line) within minutes, as management believed any worker could pick up the necessary sewing skills.

In sum, our qualitative data enable us to identify a novel mechanism—accidental gamification— to explain how quantification of work affects worker productivity. While we cannot test this proposed mechanism of accidental gamification quantitatively, our qualitative data further gives us two testable hypotheses about the conditions under which quantification of work will improve or impair productivity. Our qualitative data suggest that quantification of work is more likely to improve worker productivity in the context of simple work, where gamification is motivating because quantified metrics adequately measure the work being performed. When work is complex, by contrast, quantification is likely to hurt worker productivity because quantified metrics cannot adequately measure the multidimensional work being performed, rendering gamification demotivating. The next section describes the quantitative methods and data that we use to test these hypotheses.

**QUANTITATIVE METHODS, DATA AND MEASURES**

We test our hypotheses by exploiting a natural experiment. On October 1, 2012, plant management installed RFID measurement technology on two of the nine pants lines; two months later, on December 1, 2012, the same technology was installed on one of the three jacket lines. When the technology was installed on a line, the work of all operations within that line became quantified. We have data on daily line-level productivity from January 1, 2009, to December 31, 2014 (that is, both before and after the quantification intervention). For the nine pant lines, we also have daily operation-level productivity for a few months before and after the intervention.

This configuration enables us to investigate the effect of quantification of work on productivity in the treated lines and operations, with the nontreated lines and operations serving as a comparison group to control for time-varying trends in productivity. We can further explore heterogeneity in the effects by work complexity. A key advantage of our setting is the opportunity to exploit relatively small differences in work complexity rather than comparing vastly different kinds of work. Our empirical strategy will first exploit variation in complexity across pant and jacket lines. Figure 2 depicts this empirical strategy. We then further exploit variation in work complexity across operations within pant lines to address possible objections that pant and jacket lines could differ on dimensions other than complexity.

[Figure 2]

**Complexity Across Product Lines**

In garment manufacturing, complexity is known to vary by product. Managers and workers both told us that jackets are more complex than pants. A shop-floor supervisor said, "Because jackets are more complex overall, the individual operations being performed by workers are also more complex" (S2). An industrial engineer elaborated: "Jackets are a more complex, multifaceted product, and they have greater SMVs" (IE1). As is common in the industry, the plant rates complexity using *standard minute values* (SMVs).[6] A given SMV represents the total number of minutes allocated for workers performing at a standard level to complete a given product from start to finish; higher values correspond to more complex products. We collected data on the SMVs of different styles of pants and jackets being produced. The mean SMV of pants was

[6] SMVs are also referred to as *standard allowed minutes*, or SAMs.

28.99, with a standard deviation of 3.65; that of jackets was 88.44, with a standard deviation of 9.76. Thus, on average, producing a jacket takes almost three times longer than producing a pant. A simple t-test reveals that this difference in mean SMVs is statistically significant at the 0.01 level.[7]

We exploit this variation in complexity by first comparing how quantification affects productivity in the more complex jacket lines and in the simpler pant lines. Our dependent variable here is daily line-level productivity. Line-level productivity is measured in the plant by *percent efficiency*, a measure of productivity that accounts for the complexity of the product. For example, if the manufacturer assigns a product an SMV of 30, it expects a line to complete one unit in 30 minutes. If the line averages 50 minutes to do so, its percent efficiency is 60% (30 / 50). We have daily line-level percent-efficiency data for each of the twelve lines (9 pant lines and 3 jacket lines) between 2009 and 2014. This data was collected manually, pre- and post-treatment, by comparing the number of fully finished garments produced by a line to the daily target for that line.

**Complexity across Operations within Pant Lines**

Because we might worry that pant and jacket lines differ on dimensions other than complexity, we also exploit variation in the complexity of operations within pant lines, where operations are carefully constructed to be incapable of being broken down further. Just as the complexity of garments is rated using SMVs, so is the complexity of specific operations. Here, an SMV represents the number of minutes allocated for a qualified worker performing at a standard level

[7] If we divide the mean SMV of pants by the number of pant operations (51) and the mean SMV of jackets by the number of jacket operations (96), we see that the mean SMV of pant operations, 0.56, is still about half that of jacket operations,0.92.

to complete a given operation; higher values correspond to more complex operations. Operations assigned higher SMVs typically entail more dimensions, greater dexterity and cognitive capacity.

An industrial engineer explained: "Each worker is assigned to an operation, which should be thought of as a task that cannot be broken down any further.... And so, even in the relatively easy pant production, some tasks naturally take longer to do, have greater SMV (standard minute values) and are considered to be more complex" (IE1). For example, attaching waistbands is a complex operation, with an SMV of 0.95; that is, an operator must execute an operation every 57 seconds (including rework and rest) to achieve 100% efficiency. In contrast, attaching wash-care labels is a simple operation with an SMV of 0.19, which corresponds to performing the operation in 11.4 seconds. Though there are few complex operations within pant lines, we exploit variation in complexity across pant-line operations to investigate the effect of quantification of work on different operations. To conduct this analysis, we obtained daily operation-level productivity data within pant lines for a subset of dates in 2012. We excluded jacket lines from the operation-level analysis because we lacked pre-treatment productivity data for jacket production. For the pant lines, we had both pre- and post-treatment operation-level data.

Our dependent variable for this analysis is daily operation-level productivity. Like line-level productivity, operation-level productivity too is measured by percent efficiency, a measure that depends on both the speed of the operator and the complexity of the operation. For example, if the manufacturer rates an operation as having an SMV of 0.75, it expects a skilled operator to complete one operation in 0.75 minutes. If a particular worker takes an average of 0.50 minutes to complete that operation over the course of a shift, that worker's percent efficiency is 150% (0.75/0.50). At the end of the production line, inspectors check products and return any with deficient quality to the responsible worker. Such rework reduces a worker's measured efficiency, in that rework takes time but does not count toward items produced.

Manual records provided pre- and post-treatment operation-level productivity data for the pant lines. During the pre-treatment period, data collection at the individual level within lines was done at random, infrequent intervals, as spot checks, with the aid of a manual system in which a data-entry clerk (who had no formal authority) counted how many items workers performing specific operations had produced. Importantly, plant supervisors continued to use the manual system in the treated lines post-treatment, despite the new RFID technology given that the intervention was a pilot. Thus, to exclude the possibility that our estimated treatment effect is biased by changes in the data-collection regime, we use manually collected data to conduct both line-level and operation-level analyses.

**Empirical Strategy: Difference-in-Differences**

The data allow us to test for the effect of quantification of work using a difference-in-differences framework. The analysis thus consists of three parts: (1) the unconditional analysis, where we examine overall changes in productivity at treated lines post-treatment; (2) the analysis that conditions on the complexity of the product line; and (3) the analysis that conditions on the complexity of operations within the pant lines.

First, we estimate the effect of the quantification intervention on productivity:

$$Y_{it} = \beta_0 + \beta_1 \text{post}_{it} + \beta_2 \text{treated}_i + \beta_3 \text{posttreated}_{it} + \phi_i + \tau_t + \varepsilon_{it} \qquad (1)$$

In equation (1), $Y_{it}$ is the productivity (the percent efficiency) of line *i* on day *t*; $\text{treated}_i$ is an indicator that takes a value of 1 if line *i* is ever a treated line; and $\text{post}_{it}$ is an indicator for whether line *i* was in the post-treatment period on day *t*, regardless of whether that line was actually treated. Because of the technology's staggered introduction, this variable takes a value of 1 for lines producing pants after October 1, 2012, or producing jackets after December 1, 2012. Then, posttreated$_{it}$ is an indicator that takes a value of 1 if treated line $i$ is in the treatment period at time $t$, which is functionally equivalent to the interaction between the post and treated indicators. We run our regressions with and without line effects, denoted by $\phi_i$, and, similarly, with and without month and day-of-week fixed effects, denoted by $\tau_t$. We cluster standard errors by line to allow for errors to be correlated within lines.

As in a standard experimental difference-in-differences framework, the causal identification comes from the difference between treated and nontreated lines before and after the treatment. This yields estimates for the causal effect of the treatment, whatever the *ex-ante* differences between the lines, as long as the treatment is not correlated with other factors that also affect productivity. Although the treated lines in this case were not explicitly chosen at random, management chose those nearest to the offices of the engineers responsible for maintaining the technology. Thus the choice of treatment lines was quasi-random, made for a reason unrelated to expected trends in productivity.

As in any difference-in-difference test, our chief concern is a failure of the parallel-trends assumption: treated lines should have the same trend in productivity as nontreated lines in the absence of the treatment. To address this concern, we first plotted the weekly difference in efficiency between treated and nontreated lines both pre- and post-treatment; we found that the parallel-trends assumption was met in the pre-treatment period, where the difference in productivity between the treated and nontreated lines was flat. We also differentially estimated our main regressions with only the pre-treatment data and found no effect of quantification. Finally, we also added a linear time term to our regression specifications, restricting the sample to the pre-treatment period, and found that the coefficient on the time term was not significant.

**Complexity across Product Lines.** Next we examine how the effect of quantification of work varies by product-level complexity. To do so, we rerun equation (1) separately for lines producing jackets and pants:

$$Y_{pt} = \beta_0 + \beta_1 \text{post}_{pt} + \beta_2 \text{treated}_p + \beta_3 \text{posttreated}_{pt} + \phi_p + \tau_t + \varepsilon_{pt} \tag{2a}$$
$$Y_{jt} = \beta_0 + \beta_1 \text{post}_{jt} + \beta_2 \text{treated}_j + \beta_3 \text{posttreated}_{jt} + \phi_j + \tau_t + \varepsilon_{jt} \tag{2b}$$

where *p* denotes the nine simple pant lines, *j* denotes the three complex jacket lines, and all other variables remain the same. Hypothesis 1 concerns the direction of the coefficient on $\beta_3$ in the first equation (2a); Hypothesis 2 concerns the direction of the coefficient on $\beta_3$ in the second equation (2b).

Equations (2a) and (2b) are convenient for comparing the productivity growth of treated and nontreated lines, separately for simple and complex product lines. We may also wish to test whether the effect of the measurement technology differs in simple and complex product lines. To do so, we fully interact equation (1) with an indicator that the line produces complex jackets:

$$Y_{it} = \beta_0 + \beta_1 \text{post}_{it} + \beta_2 \text{treated}_i + \beta_3 \text{posttreated}_{it} + \beta_4 \text{complex}_i + \beta_5 post_{it} * \text{complex}_i + \beta_6 \text{treated}_i * \text{complex}_i + \beta_7 \text{posttreated}_{it} * \text{complex}_i + \phi_i + \tau_t + \varepsilon_{it} \tag{2c}$$

where $\text{complex}_i$ denotes that line *i* produces the more complex product (jackets), and all other variables remain as before. As a three-way interaction, remember that $\beta_0$ through $\beta_3$ are estimated with the simple (pant) lines. Then, $\text{posttreated}_{it}$*$\text{complex}_i$ is our coefficient of interest, estimating the difference in the effect of quantification on productivity between simple and complex product lines. As before, we run these three regressions with and without line, month, and day of week fixed effects. We cluster standard errors by line. This line-level analysis allows us to examine how quantification affected productivity depending on product-level complexity.

**Complexity across Operations Within Pant Lines.** Finally, we exploit variation in complexity across operations within pant lines. This operation-level analysis has three distinct advantages over the line-level analysis. First, pant and jacket lines could be correlated with some factor other than average complexity that yields productivity gains or losses from quantification; switching to the operation-level analysis allows us to evaluate complexity more directly. Second, there are a large number of operations, and therefore a larger number of clusters in the operation, than in line analysis, which affords greater statistical power. Third, since our measure of complexity is a continuous variable, the operation-level analysis allows us to estimate the effect of quantification at discrete intervals of complexity as well.

For these tests, our empirical strategy mirrors what we use for line complexity. We begin by distinguishing between simple and complex operations, depending on whether they are below or above the 75th percentile SMV. We run the regression using the $75^{th}$-percentile cutoff (corresponding to an SMV of 0.75) because pants operations are already relatively simple; we thus designate complex pant operations as the most complex of them.[8] The regressions take the form:

$$Y_{ist} = \beta_0 + \beta_1 \text{post}_t + \beta_2 \text{treated}_i + \beta_3 \text{posttreated}_{it} + \phi_i + \tau_t + \varepsilon_{ist} \quad (3a)$$
$$Y_{ict} = \beta_0 + \beta_1 \text{post}_t + \beta_2 \text{treated}_i + \beta_3 \text{posttreated}_{it} + \phi_i + \tau_t + \varepsilon_{ict} \quad (3b)$$

where $Y_{ist}$ and $Y_{ict}$ represent the efficiency of line *i*'s simple operation *s* or complex operation *c* at day *t*, and all other variables are as they were before. We run these regressions with and without line, month and day-of-week fixed effects and standard errors clustered by operation.

Likewise, we perform a fully interacted regression:

$$Y_{iot} = \beta_0 + \beta_1 \text{post}_t + \beta_2 \text{treated}_i + \beta_3 \text{posttreated}_{it} + \beta_4 \text{complex}_o + \beta_5 post_t * \text{complex}_o + \beta_6 \text{treated}_i * \text{complex}_o + \beta_7 \text{posttreated}_{it} * \text{complex}_o + \phi_i + \tau_t + \varepsilon_{iot} \quad (3c)$$

[8] Our results are robust to variations in this cutoff point.

where $Y_{iot}$ is the percent efficiency for line *i*'s operation *o* at month *t*. For consistency and completeness, we show results where $\text{complex}_o$ is treated as a discrete variable indicating whether the SMV rating of the operation is above the 75th percentile SMV for pant jobs, and where $\text{complex}_o$ is a continuous variable representing the SMV rating of the operation.

**RESULTS**

Table 3 shows that the treated lines were not systematically different than nontreated lines with regard to observable worker characteristics prior to installation of the measurement technology.

[Table 3]

We next investigate differences in worker characteristics between simple and complex lines, and between simple and complex operations. Workers are assigned to a single operation on a single line and rarely move. Because managers assign workers to the operations where they are needed when they finish training, we expect worker capabilities to be uncorrelated with job characteristics. Table 4 corroborates that assignments are uncorrelated with workers' skill ratings, education, prior tailoring experience and tenure.

[Table 4]

Table 5 reports summary statistics of the mean and standard error of percent efficiency (our measure of productivity) at treated and nontreated lines before and after the measurement technology was implemented, for all lines, by product line, and by operational complexity within pant lines. We also present estimates for the first-differences and difference-in-differences.

[Table 5]

Table 5, Panel 1, shows that overall productivity improved at both treated and nontreated lines; improvement was slightly greater at treated lines. Note that the mean percent efficiency is not 100% but around 60%. Thus a gain of 4% points in efficiency generally corresponds to a 7%

gain in productivity (64% / 60% = 107%). Note too that the overall differences reported in this panel include both the direct effect of quantification on pant and jacket lines and the compositional effect that places more weight on pants, accounting for nine of the twelve lines.

Panel 2a shows changes in productivity on simple product lines (two treated and seven nontreated pant lines). Productivity remained relatively flat at the nontreated lines and rose substantially at the two treated lines after introduction of the measurement technology. Panel 2b shows changes in productivity on complex product lines (one treated and two nontreated jacket lines). In contrast to the pant lines, productivity at the treated jacket lines declined even as productivity at the nontreated lines improved. These summary statistics offer preliminary support for our hypotheses by demonstrating that, when comparing the difference in productivity between treated and nontreated lines before and after the intervention, the effect of quantification on simple lines was positive and on complex lines was negative.

The third panel shows changes in productivity within the pant operations. The efficiency of simple operations (in 3a) remained approximately constant within the nontreated lines but rose by 5.1% points post-treatment at treated lines. For complex operations (in 3b), efficiency declined slightly at nontreated lines and by a larger amount at treated lines. These summary statistics offer additional support for our hypotheses by demonstrating that, as in the line-level results, the effect of quantification on simple operations was positive, while its effect on complex operations was negative. We turn next to our regression results, where we add controls and cluster standard errors.

First, we look at the unconditional regression results of investigating the effect of quantification on productivity across all lines. Note that regression estimates depart from the means presented in Table 5 because the regressions further control for line, month and day-of-week effects, and because the regression results cluster standard errors at the line level to allow for correlated observations within lines. These regressions correspond to equation (1). Table 6 presents the results.

[Table 6]

Table 6, column 1, shows the classic difference-in-differences without line or month controls. Treated lines had slightly higher productivity prior to implementation of the measurement technology, but no statistically significant growth after the technology was introduced. Column 2 introduces month and day of week fixed effects (note that the *Post* term cannot be estimated here); Column 3 introduces line fixed effects (note that the *Treated* term cannot be estimated here). Column 4 includes line, month and day-of-week fixed effects (note that the *Post* and *Treated* terms are absorbed by the fixed effects). Given that we have a balanced sample and that the treatment is largely uncorrelated with the controls, it is unsurprising that the main coefficient of interest, the difference-in-differences estimate, changes little as we introduce finer controls. This estimate is small and not statistically different from zero.

If we were to stop here, we might conclude that the measurement technology had no effect on productivity. However, the following sections will examine the heterogeneity of the treatment effect by product and operational complexity.

**Complexity across Product Lines**

We next examine the effect of quantification on simple (pants) and more complex (jackets) products, which correspond to equations 2a–2c. Table 7 presents results.

[Table 7]

Table 7, columns 1 and 2, correspond to Equation 2a and show the results of the difference-in-differences analysis of pant lines. Efficiency at treated pant lines rose an estimated

8.06% points relative to nontreated lines. Given that these lines operated at 64.2% efficiency before treatment, relative productivity improved by about 12.5%. The improvement is statistically significant. The results are substantively similar after including line, month and day-of-week fixed effects.

Table 7, columns 3 and 4, correspond to equation 2b, and show results restricting analysis to the jacket lines. Note that we now have one-third of the full sample observations because only one-third of the lines produce jackets. Compared to nontreated jacket lines, productivity at the treated jacket line dropped by 9.4% points. Again, results are similar if we include line, month, and day-of-week effects. Taken together, columns 1 through 4 suggest that the effect of the quantification treatment was positive for the simple lines but negative for the complex lines.

Table 7, columns 5 and 6, correspond to equation 2c and use triple differences to test whether the treatment effect was significantly different depending on whether a line produced pants or jackets. Before we interpret the main result, a few characteristics deserve note. First, mechanically, the triple-difference coefficient will be approximately equal to the difference between the two-way interactions in the prior columns. Second, we can estimate *Post* × *Treated* and *Post* × *Complex* in the final column because of the staggered treatment. Third, in even numbered columns, the *Treated* and *Post* terms are respectively absorbed by the line and month fixed effects. In column 6, we cannot estimate the *Complex* term and the *Treated* × *Complex* term. Now, turning to the coefficient of interest, the triple-differences model estimates that the difference in the effect of quantification between simple and complex lines is 17.5% points in the reduced model and 15.7% points after including the fixed effects. Both are statistically different from zero. Importantly, what this reveals is that the effects of quantification on simple and complex work are statistically significantly different. Thus Table 5 offers support for our hypotheses using our first measure of complexity at the product level. Also see Appendix B, which shows that these results are robust to a placebo analysis.

**Complexity across Operations within Pant Lines**

Next, using operation-level data from the pant lines, we examine how the effect of the measurement technology varies by operational complexity to address possible objections that pant and jacket lines could differ on dimensions other than complexity. These analyses correspond to Equations 3a–3c. Table 8 presents results in a setup that is very similar to Table 7.

[Table 8]

Table 8, columns 1 and 2, restrict analysis to simple pant operations (SMV $\leq 0.75$). Once again, the post and treatment terms drop out in Column 2 because they are absorbed by line, month and day-of-week fixed effects. Compared to nontreated pant operations, efficiency rose post-treatment at treated operations. The coefficients in columns 1 and 2 are significant ($p < 0.1$) and substantively large (the estimated increases in efficiency of 4.97% points and 5.36% points signify approximately 8.2% and 8.9% gains over prior efficiency of simple operations).

Table 8, columns 3 and 4, restrict analysis to complex pant operations (SMV > 0.75). Compared to nontreated pant operations, efficiency declined by 4.15% points and 3.58% points in the treated operations, before and after including fixed effects. These differences are significant, with $p < 0.05$ and $p < 0.01$ respectively; they are also substantive, representing declines in efficiency of 6.2% and 5.4% relative to the prior efficiency of complex operations.

Table 8, columns 5 and 6, include all operations and test whether the effect of quantification of simple and complex operations differs. The resulting coefficients are substantial (representing differences of more than 15% and 10% of mean productivity at pant lines respectively) and statistically significant at 1% in the reduced model and 10% in the full model.

The product-level analysis features only two products; the operation-level analysis entails numerous pant operations with varying levels of SMV ratings, allowing us to further estimate the effect of the measurement technology at different intervals of complexity. We estimate the post-treatment difference in percent efficiency of treated operations at eight discrete SMV values, rounding the SMV values of each operation to the nearest tenth. Functionally, this is equivalent to running the regression in Table 8, column 6, but with interaction terms for each discrete bin of complexity at 0.1 SMV interval. Figure 3 presents coefficient estimates and standard errors.

[Figure 3]

As Figure 3 shows, the error bars are relatively large within any one category of SMV values. Nonetheless, the results suggest a negative relationship between complexity and the effect of the measurement technology on productivity. Consistent with the earlier regression results, this finding is driven by gains in productivity among the simplest operations and losses in productivity among the most complex operations.

In sum, these results show that the effect of the quantification treatment was positive for the simple operations but negative for the complex operations, and that this difference in the effect of quantification was statistically significant. We further show that the results are robust to estimating the effect of quantification at discrete intervals of complexity. Thus this set of results supplements our analysis using product-level variation in complexity and offers further support for our hypotheses, using our second measure of complexity at the operation level.

**Alternative Explanations**

In the Results section above, we estimated the differential effect of quantification of work on workers' productivity by work complexity, using both product-level and operation-level measures of complexity. We recognize that the product-level measure raises concerns about whether pant and jacket production could be correlated with some factor other than complexity that yields productivity gains or losses from quantification. However, the subsequent operation-level analysis compares the effect of quantification of work on relatively simple and complex operations within pant lines and finds very similar results, ameliorating that concern. Further, Figure 3 goes beyond a binary categorization of simple and complex operations within pant lines to show the effect of quantification at small intervals of complexity, showing that as complexity rises, quantification has a more negative effect on productivity. These findings that the operation-level analysis mimics and extends the product-level analysis significantly narrows the range of alternative explanations since any alternative explanation simply rooted in differences in pant and jacket production would be insufficient to explain our findings.

A few alternative explanations remain. In particular, we test for four such explanations. First, we test for different compositions of workers across simple and complex lines and operations, with respect to capabilities and personalities that could drive differences in productivity in response to quantification. Second, we test for differences in motivation between workers assigned to simple and complex lines and operations, arising from career structures or status differences, that could similarly explain our results. Third, we test for differences in task interdependence in the organization of simple and complex work. Finally, we test whether mean reversion—the tendency, if a variable is extreme on its first measurement, to be closer to the mean or average on its second measurement—could account for our findings.

Appendices C, D, E and F, corresponding to the four alternative explanations, present quantitative and qualitative evidence to rule out each one. We are convinced that these alternatives are unlikely to explain the results we have presented, but recognize that they could operate in other settings.

## DISCUSSION

This paper seeks to answer the question: how and when does quantification of work affect worker productivity? We studied a garment manufacturing plant in India where an RFID quantification intervention was introduced to a subset of production lines in late 2012. We employed a full-cycle research design, conducting ethnographic fieldwork and interviews to develop theory and generate hypotheses, then testing our hypotheses using administrative data from the plant. Our qualitative data revealed that quantification can affect worker productivity via "accidental gamification," which we define as workers' inadvertent transformation of work into a game, and uncovered three conditions enabling this mechanism: (a) clarity of objective, (b) competition with oneself, and (c) interactive feedback. The qualitative data also informed our hypotheses about the differential effect of quantification of work depending on the complexity of the work performed: gamification is motivating for simple work because quantified metrics adequately measure such work, but demotivating for complex work that quantified metrics cannot adequately represent in full. Consistent with our hypotheses, we found— using our quantitative data—that productivity improved on lines and operations performing simple work and declined on those performing complex work.

### Contributions to the Quantification Literature

Our paper makes three important contributions to the study of quantification. First, we show how quantification *of work,* in itself, can affect worker behavior and performance. There is growing public, academic and governmental demand for quantification of most social phenomena. Whether in the form of efforts to incorporate scientific evidence into policy decisions, to extend market discipline to government and non-profit organizations, or to coordinate activity across geographic and cultural distances, pressures to devise quantitative measures have burgeoned in the past few decades (Porter 1995, Power 1997, Strathern 2000). Sociologists have noted the spread of quantification in a variety of domains, such as education (Espeland and Sauder 2016) and criminal justice (Brayne 2017) but have devoted less attention to quantification of the work of individual employees. Yet we are aware of a growing trend toward quantification of work in a variety of workplaces, such as among Uber drivers (Scheiber 2017), Alibaba engineers (Chen and Shepherd 2018), and Upwork freelancers (O'Donovan 2018). We hope that this paper will open the door to more research on this increasingly prevalent phenomenon.

Second, we develop the novel mechanism of accidental gamification, through which quantification of work can affect worker behavior and performance. Quantification scholars have identified three mechanisms that might explain why individuals change their behavior in response to quantification: quantification might (a) alter incentives (Sauder and Espeland 2009), (b) create new knowledge (McAfee et al 2012), or (c) give rise to greater disciplinary action (Levy 2015). We show, however, that the simple act of quantification could affect worker productivity on its own, even in the absence of formal changes in incentives, big-data analytics or new regimes of control. That is, the mere act of measurement could inadvertently change the behavior that it purports to measure. Numbers are the building blocks of many games (Edery and Mollick 2009); thus, simply affixing a number to a social phenomenon could lead workers to accidentally turn work into a game, resulting in performance changes. Analysis of our quantitative data further revealed three enablers of accidental gamification: (a) clarity of objective, (b) competition with oneself and (c) interactive feedback. We argue that when workers can identify an objective in order to "win" at the game, compare current performance to past performance and access information on where they stand, they are likely to gamify their work. As more organizations generate fine-grained data about their actions, it is important to realize that the act of quantification can produce its own consequences, distinct from the consequences of analyzing the data collected.

Third, we demonstrate that quantification of work can produce divergent effects on workers' performance depending on the complexity of the work being performed. The quantification literature tends to emphasize that quantification promotes standard reactive practices because it is essentially a vehicle of commensuration, transforming complex qualities into simple, comparable quantities that share a metric (Espeland and Sauder 2007). But we argue that some numbers represent better than others the phenomena they claim to capture. In the case of simple work, the validity of the quantified metrics is likely to be accepted by the workers, motivating gamification and productivity improvements. In the case of complex work, unidimensional quantified metrics are less likely to be seen by workers as reasonable and representative of their work, producing demotivating games and negatively impacting the work performance being measured. Thus quantification of work does not always elicit convergent reactions; sometimes it produces divergent responses. Future research should examine features other than complexity that could moderate the effect of quantification on workers' performance.

**Contributions to the Organizational Literature on Gamification**

Our paper also offers two contributions to the literature on gamification. First, this literature has shown that organizations can introduce features of game design into the workplace (Mollick and Werbach 2015, Deterding et al 2011). We instead theorize about "accidental gamification," or workers' inadvertent transformation of work into a game. We hope that further attention will be paid to situations in which workers turn work into a game, unintentionally on their own part and that of their employers.

We also consider how accidental gamification affects workers' behavior. The literature on enterprise gamification suggests that games are almost always motivating (Edery and Mollick 2009). For example, Csikszentmihalyi and Csikszentmihalyi (2000) argues that the "flow" state induced by games results in a more satisfying experience for workers, giving them a sense of mastery over their environment. We do not find that in our setting, where accidental gamification is sometimes motivating and sometimes demotivating. When work is simple and quantified metrics "make sense" to workers, the game is fun; when work is complex and quantified metrics devalue workers, the game is discouraging. Thus we have responded to a call for "future study … to evolve our understanding of the way that games can make work satisfying" (Mollick and Werbach 2015: 445).

**Contributions to the Literature on Work Characteristics and Design**

Finally, we contribute to the literature on designing work (Hackman and Oldham 1976). We demonstrate that work complexity interacts with new forms of technologically mediated quantification to affect workers' motivation and productivity. Scholars have long studied the direct effect of various work characteristics, including complexity, on workers' motivation and performance. Less attention has been paid to how work characteristics like complexity interact with various organizational initiatives to affect worker outcomes. Indeed, as some scholars have hinted, it seems reasonable to expect that the effects of a particular organizational practice will depend in part on how it interacts with job design, a topic that has received limited attention. We take a first step by investigating how work complexity moderates the effect of technologically facilitated quantification of work on worker productivity. We find that, in a context of simple work, quantification improves productivity by turning work into a motivating game; in a context of complex work, quantification hurts productivity by engendering a demotivating game that devalues workers. We hope that our study inspires more investigation of the interplay between job design and new information and communication technologies in the workplace.

**Contributions to Scholarship on Employee Monitoring**

Recent studies from diverse fields have empirically investigated the impact of monitoring interventions on worker productivity. Unlike our study, the interventions in these studies entail performance management (or quantification of work) accompanied by monitoring of the new data by managers and follow-on disciplinary action or threat of discipline. A close investigation of these studies reveals mixed evidence on whether such interventions improve or hurt worker productivity.

Some studies have found that such interventions improve productivity. Hubbard (2000) found that the use of onboard computers to monitor work improved truckers' driving. Pierce, Snow and McAfee (2015) found that algorithmic monitoring add-ons to a point-of-sale IT system enhanced productivity and reduced theft at a restaurant chain, and Staats et al. (2017) found that RFID tracking software improved handwashing compliance among nonmedical hospital workers. Bernstein and Li (2017) examined the effect on productivity of replacing performance-feedback reviews with system-generated performance data at a large service organization and also found a positive effect.

Other studies have found that similar monitoring interventions hurt productivity. Patil and Bernstein (2018) found that body-worn cameras reduced police officers' effectiveness. Similarly, Stanton and Barnes-Farrell (1996) found that workers' performance at computerized tasks suffered in response to electronic monitoring that diminished their sense of personal control. Campbell et al. (2011) found that tightly tracked business units learned more slowly than loosely monitored business units in a study of six hotel properties owned by the same firm; Bernstein

(2012) found that production lines at a telephone production facility with an open floor plan were less productive than randomly selected lines given privacy curtains.

The theory we develop could shed light into these mixed findings. We theorize that quantification of work alone sometimes improves, and at other times, impedes worker productivity depending on the complexity of the work. In line with our theory, in the studies described above, it appears that a negative effect was found in contexts where work was more complex: Stanton and Barnes-Farrell (1996) and Campbell et al. (2011) studied knowledge work, Bernstein (2012) studied complex electronic production and Patil and Bernstein (2018) studied discretion-laden policing. A positive effect was found in settings where the work was simpler: among truckers, restaurant workers, service workers and hospital staff (Hubbard 2000, Pierce et al 2015, Staats et al 2017, Bernstein and Li 2018). Thus, aside from theoretical contributions to the study of quantification, we begin to "offer a coherent frame to our own field's question of how observation affects employee performance in contemporary workplaces" in response to calls for such work in organizational theory (Bernstein 2017: 3-4).

**Limitations and Future Research**

Our study is not without limitations. First, while we develop the novel mechanism of accidental gamification and specify three conditions that enable this mechanism, we are unable to test this mechanism quantitatively. We urge future researchers to design interventions to experimentally turn on and off our proposed mechanism. Second, though the RFID intervention at the plant was implemented in 2012 and our quantitative data span the years 2009–2014, the qualitative data dates from 2014–2015; we did not observe work at the plant prior to implementation of the RFID technology. Ideally, we would have engaged in participant observation both before and after implementation, a limitation that future research could address.

More broadly, we consider the implications of quantification of work by investigating differences in job complexity in a garment plant. In an absolute sense, however, the difference in complexity is quite small. Arguably, the complex garment operations in our study are less complex than typical jobs in the universe of settings where we may wish to study the effects of quantification of productivity. For example, the work performed by doctors and surgeons is significantly more complex that the complex operations we study. We acknowledge that the differences in complexity that we study are quite small, but see that as an asset of our research design: if even these minute differences in job complexity produce dramatic differences in how quantification affects work productivity, we expect that the moderating effect of work complexity will be even more salient in other contexts. Thus the estimates from our study can be thought of as conservative. We hope that future research will study settings where the complexity of work varies more markedly, so that we can hone our understanding of quantification and work settings where quantification will have positive effects on productivity.

Finally, we had the privilege of studying a context devoid of incentives, big-data analytics and disciplinary measures. This setting offered us the necessary theoretical cleanliness to uncover the novel mechanism of accidental gamification. In other organizational settings, including some service-sector work (Doughtery and Hardy 2015) and knowledge work (Scheiber 2017), quantification is bundled with incentives and control mechanisms. It would be useful for future work to investigate the impact of these bundled interventions on worker behavior.

In sum, as quantification becomes more prevalent in the world of work, it would be valuable to enrich our theoretical understanding of how workers respond to this new organizational policy while also informing how quantification should be implemented in practice. Our study takes an important first step in this direction.

## REFERENCES


Anteby, M. 2008 "Identity incentives as an engaging form of control: Revisiting leniencies in an aeronautic plant." Organization Science, 19: 202-220.

Anteby, M., and C. Chan. 2017 "The self-fulfilling cycle of coercive surveillance." Organization Science, 29: 247-263.

Batt, R. 2015 "Electronic monitoring and control at work: What is it good for?" LERA For Libraries, 14: 1-2.

Beck, J. C., and M. Wade. 2004 *Got game. How the gamer generation is reshaping business forever*. Cambridge, MA: Harvard Business School Press.

Berman, E. P., and D. Hirschman. 2018 "The sociology of quantification: Where are we now?" Contemporary Sociology, 47: 257-266.

Bernstein, E. S. 2012 "The transparency paradox: A role for privacy in organizational learning and operational control." Administrative Science Quarterly, 57: 181–216.

Bernstein, E. S. 2017 "Making transparency transparent: The evolution of observation in management theory." Academy of Management Annals, 11: 217–266.

Bernstein, E. S., and S. Li. 2017 "Seeing where you stand: From performance feedback to performance transparency." Academy of Management Proceedings, 2017: 14752.

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

Brayne, S. 2017 "Big data surveillance: The case of policing." American Sociological Review, 82: 977-1008.

Burawoy, M. 1979 *Manufacturing Consent: Changes in the Labor Process under Monopoly Capitalism*. Berkeley, CA: University of Chicago Press.

Campbell, D., M. J. Epstein, and F. A. Martinez-Jerez. 2011 "The learning effects of monitoring." The Accounting Review, 86: 1909–1934.

Campion, M. A. 1988 "Interdisciplinary approaches to job design: A constructive replication with extensions." Journal of Applied Psychology, 73: 467.

Chen Y., and C. Shepherd. 2018 "Ding! Always-on Alibaba office app fuels backlash among Chinese workers." Business Insider. Accessed at https://www.businessinsider.com/r-ding-always-on-alibaba-office-app-fuels-backlash-among-chinese-workers-2018-8

Christin, A. 2018. "Counting clicks: Quantification and variation in web journalism in the United States and France." American Journal of Sociology, 123: 1382-1415.

Covaleski, M. A., M. W. Dirsmith, J. B. Heian, and S. Samuel. 1998 "The calculated and the avowed: Techniques of discipline and struggles over identity in Big Six public accounting firms." Administrative Science Quarterly, 43: 293-327.

Csikszentmihalyi, M., and I. Csikszentmihalyi. 2000 *Beyond Boredom and Anxiety*. San Francisco, CA: Jossey-Bass Publishers.

De Man, H. 1985 *The Psychology of Marxian Socialism*. Piscataway, New Jersey: Transaction Books.

DellaVigna, S. 2009 "Psychology and economics: Evidence from the field." Journal of Economic Literature, 47: 315 – 372.

Deterding S., D. Dixon, R. Khaled, and L. Nacke. 2011 "From game design elements to gamefulness: Defining gamification." Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments, 9-15 (ACM).

Dougherty, C., and Q. Hardy. 2015 "Managers turn to computer games, aiming for more efficient employees." New York Times. Accessed at https://www.nytimes.com/2015/03/16/technology/managers-turn-to-computer-games-aiming-for-more-efficient-employees.html

Edery, D., & E. Mollick. 2009 *Communities of Innovation: How Video Game Makers Capture Millions of Dollars of Innovation from User Communities and You Can, Too!* Upper Saddle River, NJ: Pearson Education.
Edwards, J. R., J. A. Scully, and M. D. Brtek 1999 "The measurement of work: Hierarchical representation of the multimethod job design questionnaire." Personnel Psychology, 52: 305-334.
Edwards, J. R., J. A. Scully, and M. D. Brtek 2000 "The nature and outcomes of work: A replication and extension of interdisciplinary work-design research." Journal of Applied Psychology, 85: 860.
Espeland, W. 1993 "Power, policy and paperwork: the bureaucratic representation of interests." Qualitative Sociology, 16: 297-317.
Espeland, W. N., and M. L. Stevens. 1998 "Commensuration as a social process." Annual Review of Sociology, 24: 313-343.
Espeland, W. N., and M. Sauder. 2007 "Rankings and reactivity: How public measures recreate social worlds." American Journal of Sociology, 113: 1-40.
Espeland, W. N., and M. L. Stevens. 2008 "A sociology of quantification." European Journal of Sociology, 49: 401-436.
Espeland, W. N., and M. Sauder. 2016 *Engines of Anxiety: Academic Rankings, Reputation, and Accountability*. New York, NY: Russell Sage Foundation.
Feldman, D. C., and T. W. Gainey 1997 "Patterns of telecommuting and their consequences: Framing the research agenda." Human Resource Management Review, 7: 369-388.
Fine, G. A., and K. D. Elsbach. 2000 "Ethnography and experiment in social psychological theory building: Tactics for integrating qualitative field data with quantitative lab data." Journal of Experimental Social Psychology, 36: 51-76.
Foucault, M. 1977 *Discipline and Punish: The Birth of the Prison.* Penguin Books, London.
Fourcade, M. 2011 "Cents and sensibility: economic valuation and the nature of 'nature'." American Journal of Sociology, 116: 1721-77.
Fourcade, M. and K. Healy. 2017 "Categories all the way down." Historical Social Research, 42: 286-296.
Frese, M., H. Garst and D. Fay. 2007 "Making things happen: Reciprocal relationships between work characteristics and personal initiative in a four-wave longitudinal structural equation model." Journal of Applied Psychology, 92: 1084.
Frost, and Sullivan. 2016 "Microsoft uses gamification to boost performance, skills and communication across thousands of agents." Customer Case Study. Accessed at https://www.gameffective.com/wp-content/uploads/FS_CS_GamEffective-AgentProductivity_BBR_112916_CAM-v2.pdf
George, G., M. R. Haas, and A. Pentland. 2014 "Big data and management." Academy of Management Journal, 57: 321-326.
Glaser, B. G., and A. L. Strauss. 1967 *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine Book Company.
Gray, G. C., and S. S. Silbey. 2014 "Governing inside the organization: Interpreting regulation and compliance." American Journal of Sociology, 120: 96-145.
Greve, H. R. 2003 *Organizational Learning from Performance Feedback: A Behavioral Perspective on Innovation and Change*. Cambridge, UK: Cambridge University Press.
Hackman, J. R., and G. R. Oldham. 1976 "Motivation through the design of work: Test of a theory." Organizational Behavior and Human Performance, 16: 250-279.
Hubbard, T. N. 2000 "The demand for monitoring technologies: The case of trucking." The Quarterly Journal of Economics, 115:533–560.
Humphrey, S. E., J. D. Nahrgang, and F. P. Morgeson. 2007 "Integrating motivational, social, and contextual work design features: A meta-analytic summary and theoretical extension of the work design literature." Journal of Applied Psychology, 92: 1332.

Hwang, H., and W. W. Powell. 2009 "The rationalization of charity: The influences of professionalism in the nonprofit sector." Administrative Science Quarterly, 54: 268-298.
Karunakaran, A. 2017 "Quantification in practice: Examining how analytics shape decision making in organizations." Working paper.
Keevers, L., L. Treleaven, C. Sykes and M. Darcy. 2012 "Made to measure: Taming practices with results-based accountability." Organization Studies, 33: 97-120.
Kim, M. 2019 "Unknown Unknowns: Limits of Transparency as a Means of Control" Working Paper.
Kiviat, B. 2017 "The art of deciding with data: Evidence from how employers translate credit reports into hiring decisions." Socio-Economic Review, 0:1-27.
Kohn, M. L., and C. Schooler. 1978 "The reciprocal effects of the substantive complexity of work and intellectual flexibility: A longitudinal assessment." American Journal of Sociology, 84: 24–52.
Lamont, M. 2012 "Toward a comparative sociology of valuation and evaluation." Annual Review of Sociology, 38: 201-221.
Lazear, E. P. 2000 "The power of incentives." American Economic Review, 90: 410-414.
Levy, K. E. 2015 "The contexts of control: Information, power, and truck-driving work." The Information Society, 31: 160-174.
Lupton, D. 2016 *The Quantified Self*. Cambridge, UK: John Wiley & Sons.
Lyon, D., K. Ball, and K. D. Haggerty 2012 *Routledge Handbook of Surveillance Studies*. London, UK: Routledge.
Mangalindan, J. P. 2010 "Play to win: The game-based economy." September 3. Fortune Magazine. Accessed at http://fortune.com/2010/09/03/play-to-win-the-game-based-economy/
Marx, G. T. 2016 *Windows into the Soul. Surveillance and Society in an Age of High Technology.* Chicago, IL: University of Chicago Press.
Mazmanian, M., and C. M. Beckman. 2018 "Making' your numbers: Engendering organizational control through a ritual of quantification." Organization Science, 29: 357-379.
McAfee, A., E. Brynjolfsson, T.H. Davenport, D.J. Patil, D. Barton. 2012 "Big data: the management revolution." Harvard Business Review, 90: 60-68.
Miller, P., and M. Power. 2013 "Accounting, organizing, and economizing: Connecting accounting research and organization theory." The Academy of Management Annals, 7: 557-605.
Mollick, E. and N. Rothbard. 2014 "Mandatory fun: Consent, gamification and the impact of games at work." The Wharton School Research Paper Series.
Mollick, E. and K. Werbach. 2015 "Gamification and the enterprise." In The Gameful World: Approaches, Issues, Applications: 439-538. Cambridge, MA: MIT Press.
Morgeson, F. P., and M. A. Campion. 2003 "Work design." In Handbook of Psychology: Industrial and Organizational Psychology: 423-452. Hoboken, NJ: Wiley.
Morgeson, F. P., and S. E. Humphrey. 2006 "The Work Design Questionnaire (WDQ): Developing and validating a comprehensive measure for assessing job design and the nature of work." Journal of Applied Psychology, 91: 1321.
Neff, G., and D. Nafus. 2016 *Self-Tracking*. Cambridge, MA: MIT Press.
O'Donovan, C. 2018 "This creepy time-tracking software is like having your boss watch you every second." BuzzFeed News. Accessed at https://www.buzzfeednews.com/article/carolineodonovan/upwork-freelancers-work-diary-keystrokes-screenshot
Patil, S.V. and E. Bernstein 2018 "The countervailing psychological effects of bilateral monitoring: A study of body worn cameras, autonomy, and conflict in law enforcement." Working paper.
Parker, S. K., T.D. Wall, and J. L. Cordery. 2001 "Future work design research and practice: Towards an elaborated model of work design." Journal of Occupational and Organizational Psychology, 74: 413.

Parker, S. K. 2003 “Longitudinal effects of lean production on employee outcomes and the mediating role of work characteristics.” Journal of Applied Psychology, 88: 620-634.
Parker, S. K. 2014 “Beyond motivation: Job and work design for development, health, ambidexterity, and more.” Annual Review of Psychology, 65: 661-691.
Parker, S. K., A. Van den Broeck, and D. Holman. 2017 “Work design influences: A synthesis of multilevel factors that affect the design of jobs.” Academy of Management Annals, 11: 267-308.
Pentland, A. S. 2014 “Saving big data from itself.” Scientific American, 311: 64-67.
Pierce, L., D. C. Snow, and A. McAfee. 2015 “Cleaning house: The impact of information technology monitoring on employee theft and productivity.” Management Science, 61:2299–2319.
Porter, T. M. 1995 *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life.* Princeton, NJ: Princeton University Press.
Power, M. 1997 *The Audit Society: Rituals of Verification.* Oxford: Oxford University Press.
Ranganathan, A. 2018 “The artisan and his audience: Identification with work and price-setting in a handicraft cluster in southern India.” Administrative Science Quarterly, 63: 879—909.
Roethlisberger, F. J. and W. J. Dickson.1943 *Management and the Worker*. Cambridge, MA: Harvard University Press.
Roy, D. 1952 “Quota restriction and goldbricking in a machine shop.” American Journal of Sociology, 57: 427–442.
Roy, D. 1959 “Banana time: Job satisfaction and informal interaction.” Human Organization, 18:158–168.
Sauder, M. 2008 “Interlopers and field change: The entry of US News into the field of legal education.” Administrative Science Quarterly, 53: 209-234.
Sauder, M., and W. N. Espeland. 2009 “The discipline of rankings: Tight coupling and organizational change.” American Sociological Review, 74: 63-82.
Scheiber, N. 2017 “How Uber uses psychological tricks to push its drivers’ buttons?” New York Times. Accessed at https://www.nytimes.com/interactive/2017/04/02/technology/uber-drivers-psychological-tricks.html
Sewell, G. 1998 “The discipline of teams: The control of team-based industrial work through electronic and peer surveillance.” Administrative Science Quarterly, 43: 397.
Sharkey, A. J., and P. Bromley. 2015 “Can ratings have indirect effects? Evidence from the organizational response to peers’ environmental ratings.” American Sociological Review, 80: 63-91.
Silverman, R. E. 2011 “Latest Game Theory: Mixing Work and Play.” Wall Street Journal. Accessed at https://www.wsj.com/articles/SB10001424052970204294504576615371783795248
Staats, B. R., H. Dai, D. Hofmann, and K. L. Milkman. 2017 “Motivating process compliance through individual electronic monitoring: An empirical examination of hand hygiene in healthcare.” Management Science, 63: 1563–1585.
Stanton, J. M., and J. L. Barnes-Farrell. 1996 “Effects of electronic performance monitoring on personal control, task satisfaction, and task performance.” Journal of Applied Psychology, 81: 738–745.
Strauss, A., and J. Corbin. 1990 *Basics of Qualitative Research: Grounded Theory Procedure and Techniques.* Newbury Park, CA: Sage Publications.
Van der Spiegel, J. 1995 “New information technologies and changes in work.” In Changing Nature of Work: 97-111. San Francisco, CA: Jossey-Bass.
Vough, H., and S. K. Parker. 2008 “Work design research: Still going strong.” In Handbook of Organizational Behavior: 411-427. London, UK: Sage Publications.
Xie, J. L., and G. Johns. 1995 “Job scope and stress: Can job scope be too high?” Academy of Management Journal, 38: 1288-1309.
Zichermann, G., & Cunningham, C. 2011 *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*. Sebastapol, CA: O'Reilly Media, Inc.

**FIGURES AND TABLES**

Figure 1: Technology-Enabled Quantification of Work: RFID Scanners Installed on Treated Lines


Figure 2: Study Design

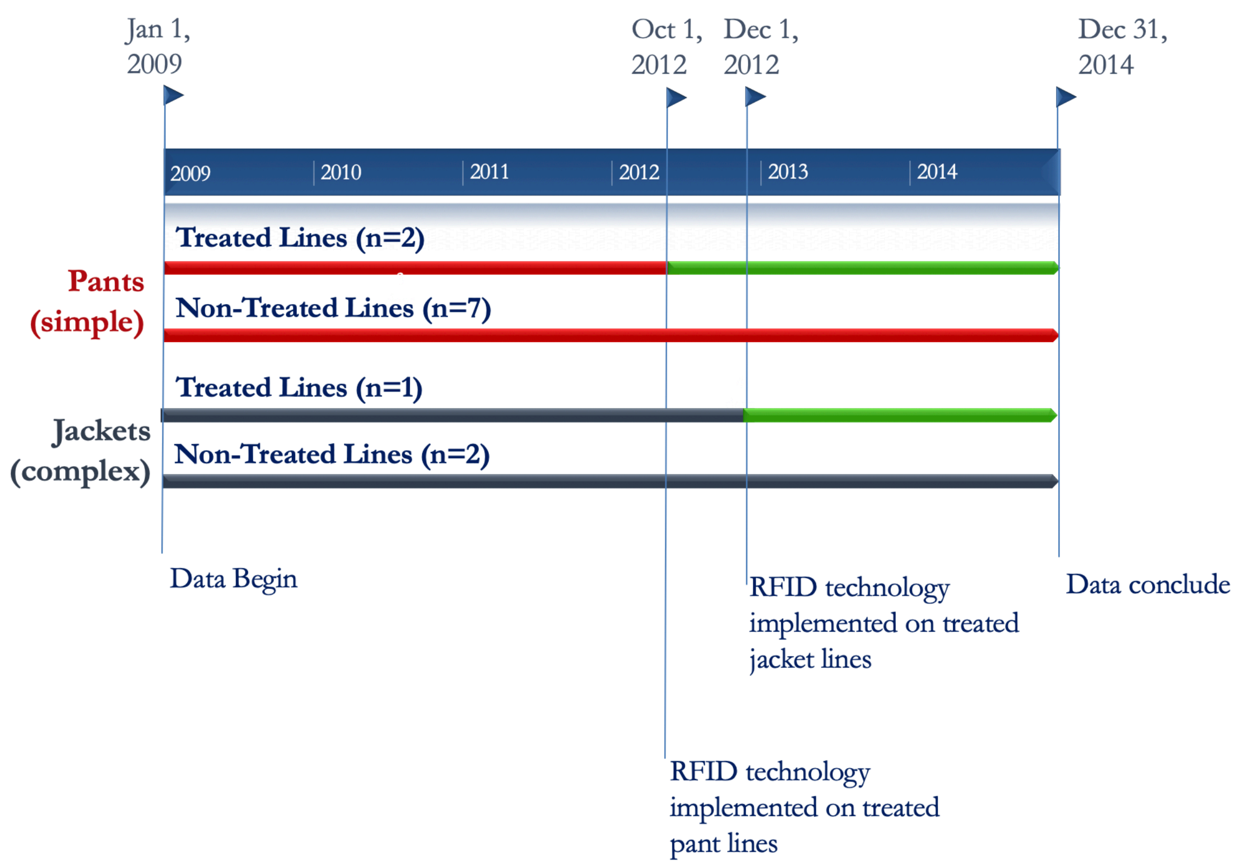
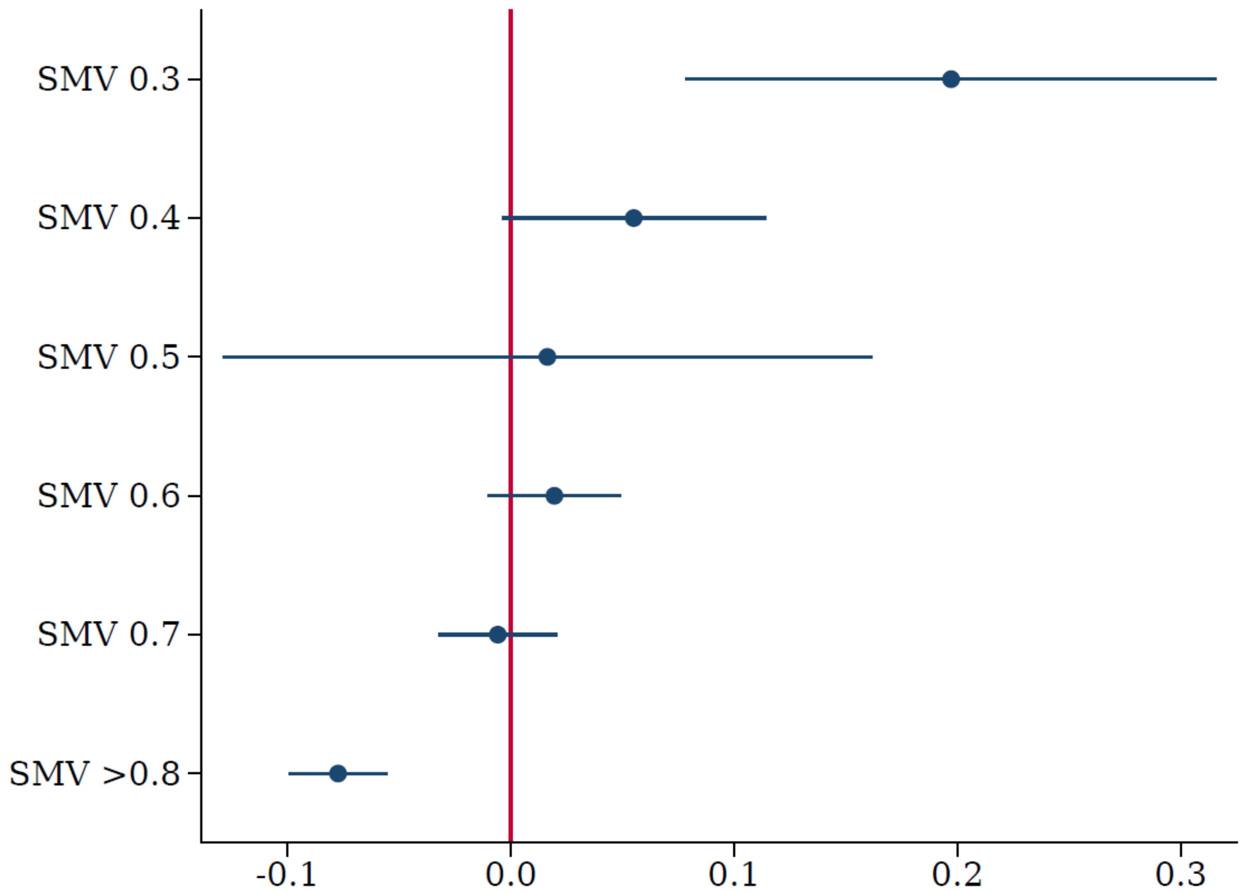


Figure 3: Difference-in-Differences Estimates of the Effect of Quantification on Productivity in Pant Lines by Operation-Complexity Bins




The figure presents difference-in-differences coefficient estimates and standard errors for the effect of treatment on efficiency, by operational complexity (rounded to the nearest 0.1 SMV).

Table 1: The Mechanism of Accidental Gamification

| | Evidence |
|---|---|
| Observations of the Mechanism of Accidental Gamification | "Workers begin looking at their efficiency because it's right there…and, before they know it, they're in a game. They get very happy when they suddenly see that they can achieve 95%....They even call us [their supervisors] over to show" (S2). |
| | "The RFID device affects workers' behavior through thought….We are not spying on them, we are not telling them to work more…and yet workers are giddy with happiness and put extra effort" (IE3). |
| **Conditions Enabling Accidental Gamification** | |
| Clarity of Objective | "The goal is to finish your piece so fast that current efficiency becomes 100%" (W5). |
| | "All operations at 100% efficiency will have a certain target.…if I work at 80% efficiency, I will be some 20 pieces short of the target.…so I need to go faster to make the target. It's simple" (W14). |
| Competition with Oneself | "The machines are very important. It helps me keep track of my production and check efficiency. I now treat the process as a competition" (W8). |
| | "I believe that my production is better in the evening because it's close to the end of the day and there isn't time left to beat my target and so there is more pressure, whereas in the morning, I think that I have the rest of the day to catch up" (W22). |
| Interactive Feedback | "The workers regularly check their production numbers and then change their behavior based on what the numbers show" (S1). |
| | "I know to check my efficiency.…I also know that if my production increases, efficiency will also increase. I can press buttons to find out other statistics too. In general, the RFID numbers don't lie" (W14). |

Table 2: How Quantification of Work Differentially Affects Simple and Complex Work

| | Evidence |
|---|---|
| **Simple Work: The Case of Pants** | |
| Definition of Work as Routine | "Pants are a classic example of mass production. You just do, do, do" (H4). |
| High Suitability of Quantity as a Metric | "A useful piece of information that I have been getting from the machine is the total number of pieces and how much I have done at a given point" (W11, working on pants). |
| Motivating Effect of Accidental Gamification | "RFID created a game…[It] made workers start focusing only on production output and everything else stopped mattering.…For my workers working on pants, that was OK. It was actually fun. They tried to do more and more pieces" (H2). |
| Positive Effect of Quantification on Productivity | "RFID drove me to produce more pieces…My target is 75 an hour, but with RFID I am able to go up to 80 an hour at times!" (W5, working on pants) |
| **Complex Work: The Case of Jackets** | |
| Definition of Work as Craft | "A jacket is a 3D garment.…It's hard to get it right. [It's not like] other products …[that] could be looked at by laying them down on a flat surface" (H4). |
| Low Suitability of Quantity as a Metric | "I am always able to deliver quality pieces, and that's what is important.…My production quantity might vary" (W20, working on jackets). |
| Demotivating Effect of Accidental Gamification | "My workers did not like it [RFID]. Sure, it's a game, but they thought they were being told by the machine 'You are not performing enough numbers.' It's not very discreet, you see—the operators have to physically scan each piece. My workers were not very keen on looking at the terminals.…[There was] the feeling of just being a number" (S4). |
| Negative Effect of Quantification on Productivity | "Workers who were used to a certain rhythm of work in a familiar, complex operation will deliver less good production now after RFID" (W21). |

Table 3: Descriptive Statistics of Workers in Treated and Nontreated Lines

| Variable | Treated Lines | Nontreated Lines | Difference |
|---|---|---|---|
| Skill rating (1 to 4) | 1.808 | 1.876 | 0.068 |
| | (0.905) | (0.900) | |
| Share: High skill (3/4) | 0.307 | 0.337 | -0.030 |
| | (0.462) | (0.478) | |
| Share: Low skill (1/2) | 0.693 | 0.663 | 0.030 |
| | (0.462) | (0.473) | |
| Education (years) | 8.627 | 8.371 | 0.246 |
| | (0.259) | (2.438) | |
| Tailoring experience (log-months) | 2.238 | 2.276 | -0.072 |
| | (1.337) | (1.151) | |
| Share female | 0.933 | 0.908 | 0.024* |
| | (0.251) | (0.289) | |
| Share from Karnataka | 0.858 | 0.837 | 0.021 |
| | (0.349) | (0.369) | |
| Age (in years) | 28.43 | 28.4 | 0.029 |
| | (6.808) | (6.484) | |
| Count of workers on same line | 245.9 | 242.1 | 3.770 |
| | (159.500) | (155.8) | |
| Tenure at plant (in years) | 1.117 | 1.013 | 0.104 |
| | (2.005) | (1.670) | |
| Number of workers | 579 | 1068 | |

* p<0.1, ** p<0.05, *** p<0.01
Standard deviations appear in parentheses.
T-tests performed; stars (if present) denote that Difference is statistically significant.

Table 4: Descriptive Statistics of Workers Performing Simple and Complex Work

Panel A: Complexity across Product Lines

| Variable | Simple lines | Complex lines | Difference |
|---|---|---|---|
| Skill rating (1 to 4) | 1.852 | 1.853 | 0.001 |
| | (0.875) | (0.917) | |
| Share: High skill (3/4) | 0.225 | 0.291 | -0.066 |
| | (0.419) | (0.456) | |
| Share: Low skill (1/2) | 0.775 | 0.708 | 0.066 |
| | (0.419) | (0.456) | |
| Education (years) | 8.457 | 8.559 | -0.102 |
| | (2.603) | (2.272) | |
| Tailoring experience (log-months) | 2.299 | 2.666 | -0.367 |
| | (1.243) | (1.005) | |
| Tenure at plant (years) | 0.335 | 0.358 | -0.023 |
| | (0.333) | (0.332) | |
| Number of workers | 586 | 1061 | |

Panel B: Complexity across Operations within Pant Lines

| Variable | Simple operations | Complex operations | Difference |
|---|---|---|---|
| Skill rating (1 to 4) | 2.471 | 2.531 | 0.061 |
| | (0.946) | (0.615) | |
| Share: High skill (3/4) | 0.549 | 0.519 | 0.030 |
| | (0.503) | (0.501) | |
| Share: Low skill (1/2) | 0.451 | 0.481 | -0.030 |
| | (0.503) | (0.501) | |
| Education | 8.471 | 8.49 | -0.019 |
| | (2.239) | (2.554) | |
| Tailoring experience (log-months) | 1.277 | 1.47 | -0.193 |
| | (1.163) | (1.262) | |
| Tenure at plant (years) | 0.331 | 0.293 | 0.038 |
| | (0.333) | (0.285) | |
| Number of workers^ | 111 | 100 | |

* p<0.1, ** p<0.05, *** p<0.01

Standard deviations appear in parentheses.

T-tests performed; stars (if present) denote that Difference is statistically significant.

^We do not have descriptive data for 77 workers in Panel B (operation-level analysis relies on data from 288 workers but we have descriptive data for only 211 workers).

Data on education and tailoring experience come from a reduced sample of workers who entered their jobs in the study period: 510 in Panel A, 115 in Panel B.

Table 5: Summary Statistics: Effect of Quantification on Productivity

| | Before RFID Implemented | | After RFID Implemented | | Difference: After−Before | |
|---|---|---|---|---|---|---|
| | Mean | SE | Mean | SE | Mean | SE |
| **Panel 1: Unconditional Analysis** | | | | | | |
| Across All Lines | | | | | | |
| Treated | 65.9% | (0.2) | 70.0% | (0.2) | 4.2% | (0.3) |
| Nontreated | 59.0% | (0.2) | 60.6% | (0.2) | 1.6% | (0.3) |
| Treated−Nontreated | 6.8% | (0.3) | 9.4% | (0.3) | 2.6% | (0.4) |
| **Panel 2: Analysis by Product** | | | | | | |
| A. Simple Product Lines (Pants) | | | | | | |
| Treated | 64.2% | (0.2) | 73.0% | (0.2) | 8.8% | (0.3) |
| Nontreated | 59.4% | (0.2) | 60.1% | (0.3) | 0.7% | (0.3) |
| Treated−Nontreated | 4.8% | (0.3) | 12.9% | (0.4) | 8.1% | (0.5) |
| B. Complex Product Lines (Jackets) | | | | | | |
| Treated | 68.8% | (0.4) | 63.9% | (0.4) | -4.9% | (0.6) |
| Nontreated | 57.5% | (0.4) | 62.1% | (0.3) | 4.5% | (0.5) |
| Treated−Nontreated | 11.2% | (0.5) | 1.8% | (0.6) | -9.4% | (0.8) |
| **Panel 3: Analysis by Operation** | | | | | | |
| A. Simple Operations | | | | | | |
| Treated | 60.0% | (1.2) | 65.1% | (1.2) | 5.1% | (1.8) |
| Nontreated | 57.1% | (1.2) | 57.2% | (0.7) | 0.1% | (0.1) |
| Treated-Nontreated | 2.9% | (1.8) | 7.9% | (1.5) | 5.0% | (2.4) |
| B. Complex Operations | | | | | | |
| Treated | 71.6% | (1.4) | 66.4% | (1.3) | -5.2% | (1.9) |
| Nontreated | 65.4% | (1.0) | 64.4% | (0.7) | -1.0% | (1.2) |
| Treated-Nontreated | 6.2% | (1.7) | 2.0% | (1.5) | -4.2% | (2.2) |

Cells represent mean percent efficiency at the line-day level (1, 2a, and 2b) and operation-day level (3a and 3b). Analysis 2a is for simple product lines producing pants, 2b is for complex product lines producing jackets, 3a is for relatively simple operations on the pant lines, and 3b is for relatively complex operations on the pant lines.

Table 6: Difference-in-Differences Regression for the Effect of Quantification on Productivity across All Lines

| | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| Post | 1.60 | | 1.32 | |
| | (2.08) | | (1.90) | |
| Treated | 6.82** | 6.85** | | |
| | (2.37) | (2.31) | | |
| Post × Treated | 2.57 | 2.65 | 2.83 | 3.00 |
| | (4.46) | (4.30) | (4.40) | (4.21) |
| Line FE | No | No | Yes | Yes |
| Month-Year FE | No | Yes | No | Yes |
| Day of Week FE | No | Yes | No | Yes |
| Observations | 16,436 | 16,436 | 16,436 | 16,436 |

* p<0.1, ** p<0.05, *** p<0.01
Observations are at the line-date level. All estimates are from OLS models.
DV: One point is 1 percent efficiency.
Post = 1 after RFID implemented.
Treated = 1 for lines that received RFID.
Standard errors clustered by line appear in parentheses.

Table 7: Difference-in-Differences Regression for the Effect of Quantification on Productivity by Product Complexity

| | Simple Lines (Pants) | | Complex Lines (Jackets) | | All Lines (Pants & Jackets) | |
|---|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
| Post | 0.73 | | 4.52 | | 0.73 | |
| | (2.50) | | (2.86) | | (2.46) | |
| Treated | 4.80* | | 11.20*** | | 4.80* | |
| | (2.52) | | (0.31) | | (2.48) | |
| Post × Treated | 8.06** | 7.76** | -9.42* | -7.79** | 8.06** | 7.95** |
| | (2.60) | (2.68) | (2.86) | (1.38) | (2.64) | (2.48) |
| Complex | | | | | -1.86 | |
| | | | | | (2.11) | |
| Post × Complex | | | | | 3.8 | 2.27 |
| | | | | | (3.46) | (2.96) |
| Treated × Complex | | | | | 6.42** | |
| | | | | | (2.50) | |
| Post × Treated × Complex | | | | | -17.5*** | -15.7*** |
| | | | | | (3.59) | (2.90) |
| Line FE | No | Yes | No | Yes | No | Yes |
| Month-Year FE | No | Yes | No | Yes | No | Yes |
| Day of Week FE | No | Yes | No | Yes | No | Yes |
| Observations | 12,137 | 12,137 | 4,299 | 4,299 | 16,436 | 16,436 |

* p<0.1, ** p<0.05, *** p<0.01
Observations are at the line-date level. All estimates are from OLS models.
DV: One point is 1 percent efficiency.
Post = 1 after RFID implemented.
Treated = 1 for lines that received RFID.
Complex = 1 for jacket lines.
Standard errors clustered by line appear in parentheses.

Table 8: Difference-in-Differences Regression for the Effect of Quantification on Productivity, by Operational Complexity in Pant Lines in 2012

| | Simple Operations (Within Pants) | | Complex Operations (Within Pants) | | All Operations (Within Pants) | |
|---|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
| Post | 0.1 | | -1.04 | | 0.1 | |
| | (0.87) | | (1.66) | | (0.87) | |
| Treated | 2.93 | | 6.21* | | 2.93 | |
| | (2.63) | | (3.17) | | (2.62) | |
| Post × Treated | 4.97* | 5.36* | -4.15** | -3.58*** | 4.97* | 5.27* |
| | (2.77) | (3.07) | (1.52) | (0.88) | (2.76) | (2.95) |
| Complex | | | | | 8.32 | 11.3 |
| | | | | | (7.33) | (7.16) |
| Post × Complex | | | | | -1.14 | -1.12 |
| | | | | | (1.80) | (1.77) |
| Treated × Complex | | | | | 3.29 | 1.82 |
| | | | | | (3.98) | (3.84) |
| Post × Treated × Complex | | | | | -9.13*** | -6.36* |
| | | | | | (3.11) | (3.58) |
| Line FE | No | Yes | No | Yes | No | Yes |
| Month FE | No | Yes | No | Yes | No | Yes |
| Day of the Week FE | No | Yes | No | Yes | No | Yes |
| Observations | 3,745 | 3,745 | 1,618 | 1,618 | 5,363 | 5,363 |

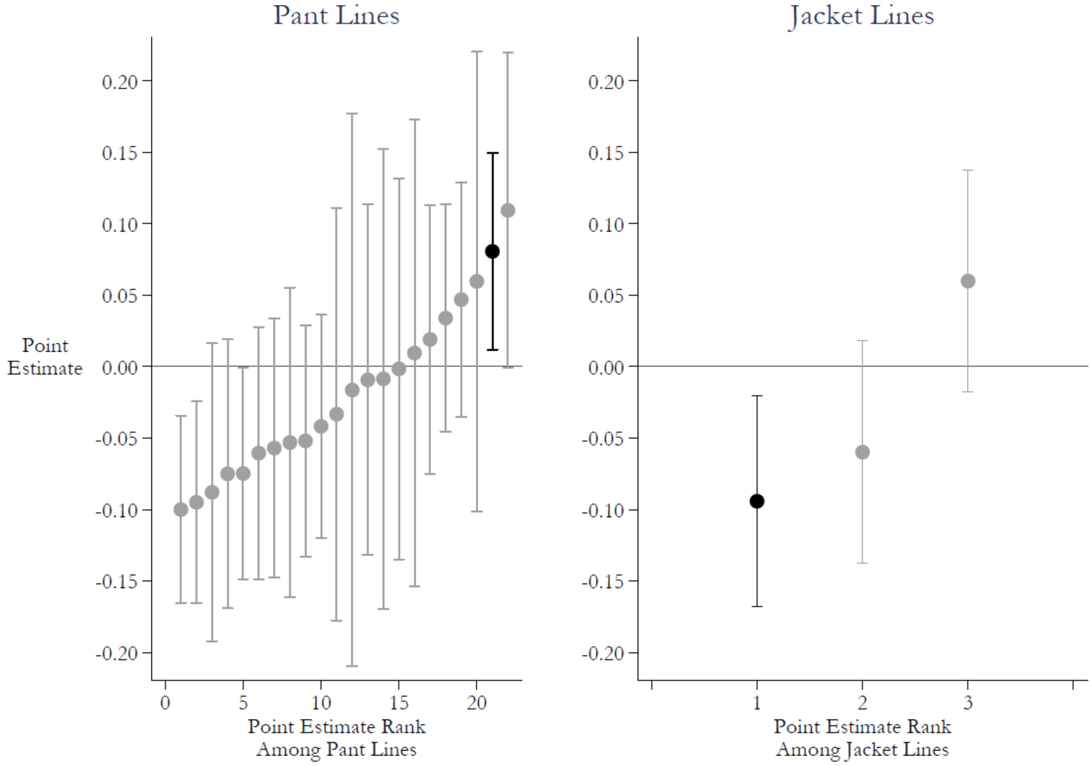
* p<0.1, ** p<0.05, *** p<0.01
Observations are at the operation-date level. All estimates are from OLS models.
DV: One point is 1 percent efficiency.
Post = 1 after RFID implemented.
Treated = 1 for lines that received RFID.
Complex = 1 for SMV greater than 0.75.
Standard errors clustered by operation appear in parentheses.

**Appendix A: Interviewees**

Table A1: Interviewee Sample (n=40)

| IDs | Sex | Group | Product | Role and respective operation |
|---|---|---|---|---|
| W1, W2, W3, W4 | F | Nontreated | Pants | Tailors: Waistband blind hem, side seam (two workers), fly top stitch |
| W5, W6, W7, W8, W9, W10, W11, W12, W13 | F | Treated | Pants | Tailors: Left fly attach, waistband label attach, trimming, coin pocket, assembly loading, fly blind hem, waistband pick ready, back pocket bag close, inseam |
| W14, W15, W17 | F | Nontreated | Jackets | Tailors: Pick stitch, breast pocket attach, neck marking |
| W16 | M | Nontreated | Jackets | Tailor: Shoulder seam |
| W17, W18, W19, W20, W21, W23, W24, W25, W26 | F | Treated | Jackets | Tailors: Neck marking, sleeve attach, front dart pressing, sleeve lining, lapel step seam, sleeve headroll attach, internal pocket sew, arm hole, under collar zig zag |
| W22 | M | Treated | Jackets | Tailor: Lapel step seam |
| S1 | F | Nontreated | Pants | Supervisor |
| S2 | F | Nontreated | Jackets | Supervisor |
| S3 | M | Treated | Pants | Supervisor |
| S4 | M | Treated | Jackets | Supervisor |
| IE1, IE2, IE3, IE4 | M | | | Industrial engineers |
| H1 | M | | | HR Head |
| H2 | M | | | Pants production lead |
| H3 | F | | | Jackets production lead |
| H4 | M | | | Head of production |
| H5 | M | | | Head of factory |
| H6 | M | | | CEO of firm |

**Appendix B: Placebo Analysis with Truly Nontreated Lines Represented as Treated Lines**




There are nine pant lines, two of which are truly treated; we first represent these two lines as non-treated. Then, from the remaining seven truly non-treated lines, we sequentially indicate every combination of two lines as treated, giving us a total of 21 placebo treatments (7 choose 2). Finally, we determine how often this procedure reproduces the results observed when we properly designate the truly treated lines.

We do the same for the jacket lines. Note that there are 3 jacket lines, one of which is treated; the same procedure yields 2 placebo treatments for jacket lines. Since the placebo interventions are uncorrelated with the true intervention, the same result should occur no more often than we would expect by statistical chance. Alternatively, idiosyncratic line-specific time trends, coupled with a finite sample of treated lines, would generate a large number of false positives among the placebos. We rank-order the placebo and nonplacebo coefficients for the *Post x Treated* term and present these with their 95% confidence intervals.

Note that the solid markers denote the nonplacebo tests. For the pant lines, none of the 21 placebo treatments generates a positive and statistically significant result with $p < 0.05$, although one placebo comes close ($p = 0.053$). This placebo treatment features the two nontreated lines that exhibited the largest growth in productivity during the treatment period, excluding the two treatment lines. Two of the 21 combinations of placebo treatments yield a statistically significant *decline* relative to the other lines. For the jacket lines, the effects of the placebo treatments are not significantly different from zero. We conclude that the line placebos do not yield false positives at a greater rate than we would expect by statistical chance.

## Appendix C: Worker Capabilities and Personalities

Table C1: Qualitative Data Ruling Out Differential Capabilities and Personalities as Alternative Explanations

| Explanations | Evidence |
| --- | --- |
| **Differential Capabilities**: Assignment of more capable workers to complex work than to simple work, such that quantification hurts productivity of complex work but not of simple work | "We don't make any distinctions between the workers in assigning them to lines. … It's not like skilled workers were put on the jacket lines to make the production better. … [Also], this [effect of RFID] does not depend on the workers.... It's not like skilled workers respond [to RFID] badly…and other kinds of workers respond better." – H2<br><br>"Jackets do not have more skilled operators than pants. RFID will be useful only if it is utilized well, and if it is installed in all lines and the data is monitored well." – IE2<br><br>"RFID does not depend on the worker; it just throws data at you irrespective of the worker [skilled or highly skilled]. How the data is used is very critical." – H3 |
| **Differential Personalities**: Assignment of more conscientious workers, for example, to simple work than to complex work, such that quantification improves productivity of simple work but not of complex work | "Every line has some workers who will be able to move production without anyone supervising them. Jackets Line 1 has some workers of the same nature – the kind that if told something has to be done, they will do it. So does Pants Line 1 and all other lines at the factory. This is a characteristic that cannot be forced on a person; they either have it or they don't …and it is not something we can easily identify. … Both before and after RFID, these workers will have a higher degree of self-interest." – H2 |

## Appendix D: Career Structures and Status Differences

Table D1: Qualitative Data Ruling out Differences in Motivation Due to Career Structures or Status Differences as Alternative Explanations

| Explanations | Evidence |
|---|---|
| **Career Structures:** More opportunities for workers who perform simple work to be promoted to complex and other higher-order tasks, causing these workers to be motivated to respond positively to quantification | "Workers don't generally move…I will be completing 3 years in the factory this October; I have been doing the same operation from the time I joined." – W4<br><br>"Promotions are extremely rare occurrences." – H2<br><br>"As a factory, firing workers…it almost never happens. We don't believe that by the [threat of] firing, workers will get their job done." – H6 |
| **Status Differences:** Higher status attributed to workers who perform complex work, causing these workers to be less motivated to cooperate or to respond favorably to quantification | Interviewer: "Where would workers prefer to work, jackets or pants?" Interviewee: "Both, it's all the same." – S4<br><br>"All operations are good and equal." – W12<br><br>"No matter what your operation, workers will have to continue doing their production.…Everyone is treated the same by supervisors and their peers." – W14<br><br>"Our operators are very flexible and good. No one gives attitude to other workers because they think they are better.…This is important." – IE4 |

## Appendix E: Task Interdependence

Table E1: Qualitative Data Ruling out Differences in Task Interdependence as an Alternative Explanation

| Explanation | Evidence |
|---|---|
| **Differential Interdependence**: Greater interdependence among workers who perform complex work than simple work, such that quantification affects this interdependence and hurts productivity of complex work | "As the SMV increases, there will be more operators assigned to that operation. Essentially what this means is that, when the operation is complex, the target reduces– and therefore, in order to meet the line target (which is about 60), there might be 2 operators doing 30 pieces each. …. But for these multiple operators working on the same operation, they always worked independently even before RFID … and with RFID it's no different…. They have to focus on their individual target; there is no combined target, no dependence on one another." – S3<br><br>"The sequence of the operation, and whether or not there is one person doing the operation or two, does not matter for RFID.… The respective job code is fed into all the RFID machines, and the worker just has to focus on their own task." – IE4<br><br>" I do the same operation as [W23]. With the RFID machines, you are right, operators solely focus on maintaining their own production numbers and efficiency percentage. But what you don't realize is that this was always true.…There were cases before also where there were two workers assigned to a single operation, where one worker was performing well and the other was struggling.…The better-performing worker would not help the other clear the pile.…It's not that, before RFID, they worked as a team and now they don't, and that's why their productivity has dropped." – W22 |

## Appendix F: Mean Reversion

Table F1: Mean Productivity of the Treated and Nontreated Complex Product Lines in the Pre-Treatment Period

| Month-Year | Treated Lines | Nontreated Lines | Difference |
|---|---|---|---|
| December, 2011 | 0.535 | 0.501 | 0.034 |
| January, 2012 | 0.508 | 0.512 | -0.004 |
| February, 2012 | 0.577 | 0.515 | 0.062 |
| March, 2012 | 0.63 | 0.602 | 0.028 |
| April, 2012 | 0.65 | 0.678 | -0.028 |
| May, 2012 | 0.629 | 0.684 | -0.055 |
| June, 2012 | 0.608 | 0.646 | -0.038 |
| July, 2012 | 0.712 | 0.67 | 0.042 |
| August, 2012 | 0.741 | 0.531 | 0.21 |
| September, 2012 | 0.72 | 0.664 | 0.056 |
| October, 2012 | 0.617 | 0.578 | 0.039 |
| November, 2012 | 0.623 | 0.577 | 0.046 |

*Note*. The month-years shown represent one year of pre-treatment data. As is clear from this table, the difference between treated and nontreated lines is small, sometimes positive and sometimes negative, and does not follow a consistent pattern. These data rule out the possibility that treated and nontreated lines were converging in productivity in the year preceding the intervention.

Table F2: Difference-in-Differences Regression for the Effect of Quantification on Productivity by Product Complexity (recapitulating Table 7 with the inclusion of terms for mean reversion)

| | Simple (Pants) | Complex (Jackets) | All Lines (Pants & Jackets) | | |
|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| Post × Treated | 6.501*** | -5.278*** | 6.500*** | 7.900*** | 6.500*** |
| | (1.519) | (1.613) | (0.377) | (2.465) | (1.533) |
| Post × Complex | | | 0.977 | 1.743 | 0.629 |
| | | | (2.15) | (2.61) | (2.35) |
| Post × Treated × Complex | | | -11.75*** | -15.38*** | -11.60*** |
| | | | (2.17) | (2.75) | (2.44) |
| Lag term | 0.489*** | 0.396*** | 0.466*** | | 0.465*** |
| | (0.069) | (0.018) | (0.057) | | (0.054) |
| Time trend | -0.0439 | 0.0455 | | -0.187 | -0.0421 |
| (in days / 100) | (0.109) | (0.241) | | (0.220) | (0.113) |
| Trend × post × treated | -0.282 | -1.13*** | | 0.053 | 0.424* |
| | (0.207) | (0.241) | | (0.044) | (0.201) |
| Line FE | Yes | Yes | Yes | Yes | Yes |
| Month-Year FE | Yes | Yes | Yes | No | No |
| Day of Week FE | Yes | Yes | Yes | No | No |
| Observations | 12,129 | 4,294 | 16,423 | 16,423 | 16,423 |

* p<0.1, ** p <0.05, *** p<0.01
Observations are at the line-date level. All estimates are from OLS models.
DV: One point is 1 percent efficiency.
Post = 1 after RFID implemented.
Treated = 1 for lines that received RFID.
Complex = 1 for jacket lines.
Lag term is a one-day lag.
Time trend takes the value 1 for day 1 of the observation window, 2 for day 2, upto day 1851.
Standard errors clustered by line appear in parentheses.