HARVARD | BUSINESS | SCHOOL

In Search of Organizational Alignment Using a 360-Degree Assessment System: A Field Experiment in a Retail Chain

Carolyn Deller
Susanna Gallani
Tatiana Sandino

Working Paper 18-069

In Search of Organizational Alignment Using a 360-Degree Assessment System: A Field Experiment in a Retail Chain

Carolyn Deller
University of Pennsylvania

Susanna Gallani
Harvard Business School

Tatiana Sandino
Harvard Business School

Working Paper 18-069

In Search of Organizational Alignment Using a 360-Degree Assessment System:

A Field Experiment in a Retail Chain

Carolyn Deller
The Wharton School
University of Pennsylvania

Susanna Gallani
Harvard Business School

Tatiana Sandino
Harvard Business School

March 2019

Abstract:

We analyze the effects of a field experiment introducing a values-based 360-degree assessment system at an Indian retailer. The director intended to encourage store managers, rewarded based on high-powered incentives linked to financial results, to behave according to the organization's long term values and goals. Surprisingly, we find that the intervention drove even higher effort on performance associated with pre-existing monetary incentives, but, on average, did not affect nonfinancial performance dimensions linked to long term goals. We integrate our statistical results with qualitative information from interviews, which highlighted the importance of reinforcing the organizational goals' message and providing support for their attainment. We also show more favorable effects for stores with tenured managers and higher availability of inventory (a proxy for support). Our findings highlight important factors for successful implementations of 360-degree systems as complements to explicit incentives. Finally, we share some lessons learned with respect to performing field experiments.

Keywords: 360-degree assessments, values, implicit incentives, field experiment, performance evaluation.
JEL codes: M12, M40
Data availability: The data used in this project was provided to the authors on a proprietary basis and cannot be shared without express consent of the organization's legal representatives.
Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sector.

We thank Markus Arnold, Tony Davila, John Harry Evans, Mario Schabus, and participants at the 2018 AAA Annual Meeting, 2018 Emerging Management Accounting Scholars Symposium, 2018 Global Management Accounting Research Symposium, 2018 Management Accounting Section Midyear Meeting, and the Harvard Business School 2017 Brown Bag Series for their useful comments and suggestions. We thank Kyle Thomas for his excellent assistance at the research site, as well as in the data collection and analyses, Mayuresh Kumar and Gunjan Kapil for their research assistance at the research site, Trang Nguyen for research assistance in the later stages of the project, and Andrew Marder and Xiang Ao for helpful advice on econometric issues. We also thank the mobile phone retail company where we conducted our study for collaborating with us on this field experiment and providing us with access to data. All errors are our own.

1. INTRODUCTION

In this paper, we study the implementation of a *values-based 360-degree performance assessment system* designed to communicate and reinforce an organization's goals and values. We do so in a setting in which employees received high-powered incentives explicitly linked to financial results (hereafter, high-powered explicit incentives). High-powered explicit incentives can motivate employees to behave entrepreneurially, but can also lead to an excessive focus on short-term, individual-wealth-maximizing activities at the expense of longer-term organizational sustainability and growth (Holmstrom and Milgrom 1991; Baker, Gibbons, and Murphy 1994). Organizations may be able to correct such distortions while still fostering entrepreneurialism by also promoting behaviors aligned with long-term goals. We study the efficacy with which a 360-degree system, implemented with this objective, led to changes in performance and behavior that were actionable within a relatively short timeframe.

360-degree performance assessment systems are designed to provide a complete evaluation from multiple appraisers: oneself and one's subordinates, supervisors, peers, customers, and so on (London and Smither 1995). They are typically used to evaluate not only performance, but also behaviors—leadership, communication, participation, and teamwork—that are otherwise difficult to measure (London and Beatty 1993). By highlighting specific performance dimensions and desired behaviors, 360-degree systems can communicate organizational priorities. In fact, several scholars have emphasized their role in shaping organizational culture and introducing or reinforcing its associated values (London and Beatty 1993; London and Smither 1995).

Studies have examined the performance effects of 360-degree systems by comparing initial ratings to subsequent ratings (e.g., Smither, London, and Reilly 2005), but the effects are unlikely to be uniform. Consequently, scholars have advocated for research on the conditions under which

360-degree systems are more or less likely to be effective (Smither, London, and Reilly 2005). In our study, the system is not explicitly linked to any rewards (which appears to be typical) and two institutional features are particularly important: The system is (a) centered on organizational goals and values, for which reason we call it a *values-based* 360-degree performance assessment system, and (b) designed to complement existing, high-powered explicit incentives.

A values-based 360-degree performance assessment system communicates what behaviors are valued in the organization. This is likely to provide implicit incentives, especially to employees interested in pursuing a career within the organization. For employees naturally aligned with company values, such systems could reinforce intrinsic motivation and encourage collaboration with others in pursuit of the organization's success (van Knippenberg 2000; Akerlof and Kranton 2005). Conversely, in the presence of high-powered explicit incentives, a 360-degree system not linked to pay may be ignored and employees may not embrace long-term goals. Studying the effectiveness of a values-based 360-degree system in such a context can be valuable because many organizations provide high-powered explicit incentive contracts—allowing employees to pursue financial success with significant freedom—but still wish to direct employees to behave in accordance with the company's values and long-term goals. Our aim is to offer insights on whether 360-degree systems can help achieve both short-term and long-term organizational objectives.

For this study, we partnered with a growing retail chain, in a major city in India that had traditionally compensated store employees with high-powered explicit sales incentives. While generally congruent with the firm's overall goals—growth and market share—these incentives had also introduced selling behaviors (such as deceiving customers and gaming the system) that were detrimental to those goals in the long term. When the company was sufficiently small, the managing director could personally monitor and shape the behaviors of employees through

informal communication of her vision and values. However, as the company scaled up, she needed a more formal system to communicate and uphold the company's vision and values.

Management therefore introduced a values-based 360-degree assessment system in tandem with a formalized vision and core values. The 360-degree system was centered on the store manager and was not tied to any incentives. Appraisals were based on surveys completed by store managers themselves, their staff, and their supervisors. The survey questions were arranged by the company's four core values, capturing the behaviors that senior management wanted. The goal was to communicate those values and provide feedback and coaching to store managers, who were expected to lead the employees by example. Existing monetary incentives were kept.

To test its effectiveness, the 360-degree system (as well as the core values and vision) was introduced in a randomly selected treatment group of half of the chain's stores. We compare the performance of treated stores with that of the control stores on various dimensions before and after the introduction of the 360-degree system, but *before* any feedback was provided to store managers at treated stores.[1] Thus, rather than documenting the feedback effects of a 360-degree values-based system, we examine its motivational effects; that is, changes in behavior driven by employees learning about, and expecting to be evaluated on, the company's core values. We focus our analysis on company-values–related metrics that could be influenced by the store staff within a short timeframe, potentially making a difference in a few days. We supplement our quantitative analyses with qualitative insights gained from interviews conducted after the implementation.

We find that the 360-degree system positively impacts measures of financial performance rewarded under the existing incentive scheme, relative to a control group; specifically, sales and gross profits are 23% and 30% higher, respectively, in treatment stores. However, we find no

[1] We provide justification for this design choice in section 3.3.

significant effects attributable to the intervention on measures associated with long-term goals and values.[2] In section 4.2.2, we explain that this lack of results may be explained by the effects not being large enough to be detected given the power of our analyses.

Our qualitative interviews suggest that the lack of improvement in performance dimensions related to longer-term organizational goals and values may also have been due to (a) low recall and understanding of the core values, and an interpretation of the initiative as an encouragement to work harder; (b) perceived inability to act according to some of the core values (in particular, employees said that they did not know how to implement the core value of "giving more value" to customers, due to a lack of tools to manage customer relations); and (c) perceptions by some that they lacked support (some referred to inventory shortages and not being paid on time).

Based on these qualitative insights, we performed supplemental analyses to better understand the circumstances in which the system may have been more or less effective. Specifically, we examined whether store manager tenure (a proxy for employees' ability to implement the core values) moderated the effect of the intervention. We find that, in stores where the manager had been with the firm longer, the intervention was associated with an increase in net promoter score (a proxy for good customer service). To examine the employees' concern that they lacked support, we examined whether stores with greater days sales in inventory (DSI) performed better after the implementation of the system. Stores with greater availability of inventory did better at bundling products through the use of promotions (a key way to provide customer value) compared to stores with lower DSI, evidence that greater support might lead to better implementation of the core values in everyday operations.

[2] Note that a favorable effect on sales and gross profits without a deterioration of measures associated with long-term goals and values (measures capturing customer service, honesty, etc.) can be viewed positively. It suggests that the system led to an increase in effort rather than to short-term improvements achieved through cutting corners.

We also share some lessons that might be helpful to researchers designing their own field experiments. For instance, we reflect on the trade-off between the strength of treatment and the risk of potential contamination, the value of conducting qualitative interviews, and the need to be cognizant of cross-country cultural differences.

We offer several contributions to the literature and to the practice of management accounting. First, we extend academic knowledge on 360-degree performance assessment systems by examining their efficacy in a common but previously unexamined setting; that is, an organization with high-powered explicit incentive contracts seeking to reinforce organizational goals and values and to shift employee mindsets towards longer-term organizational objectives. Furthermore, rather than examining the effects of the system on *perceptions* of performance captured by 360-degree surveys—that is, changes in performance ratings over time—we examine its effects on more objective performance measures related to the company's core values, measures largely free of biases—such as leniency bias—typical of 360-degree performance assessment systems.

Second, we contribute to the literature on mechanisms to improve performance without explicit monetary incentives. We examine a values-based 360-degree system as a particular mechanism in our setting, finding a positive impact on financial performance in the absence of additional tangible rewards. Prior work has examined other mechanisms that may improve performance, such as implicit promotion incentives (Campbell 2008), relative performance information (Blanes i Vidal and Nossol 2011), and recognition (Bradler, Dur, Neckermann, and Non 2016).

Third, we contribute to the literature on using multiple performance measures (e.g., Ittner, Larcker, and Randall 2003; Hall 2008), here in the context of a 360-degree assessment system. Our results are consistent with the conjecture that the introduction of new goals and values and the extensiveness of the 360-degree survey may cause employees to focus on short-term, already

rewarded and well-understood behaviors at the expense of longer-term behaviors only indirectly related to financial incentives. We highlight factors that may be important when asking more of employees; for instance, adequate resources, training, and managerial commitment.

Finally, our paper adds to the accounting literature on field-based research (e.g., Bloomfield, Nelson, and Soltes 2016; Deller 2018). We share various tips for conducting field experiments, expanding upon those originally appearing in List (2011) and reiterated in Floyd and List (2016).

The rest of the paper is organized as follows: Section 2 reviews the literature. Section 3 describes our field setting and field experiment. Section 4 presents our research design, empirical tests, and results. Section 5 contains qualitative insights from interviews and supplementary analyses. Section 6 speaks to lessons learned during this field experiment, and Section 7 concludes.

2. LITERATURE REVIEW

Organizations that motivate employees with explicit incentive contracts often use subjective performance assessments to overcome the limitation that such contracts neither capture nor promote all behaviors needed to create value (Baker et al. 1994; Gibbs et al. 2004). While these subjective assessments are often performed solely by the supervisor, some organizations use 360-degree performance assessment systems for a more complete view of performance on important dimensions that the supervisor may not be able to observe directly, such as teamwork and communication. Moreover 360-degree systems can increase employee awareness of and alignment with strategic goals and values (London and Beatty 1993).

2.1 Effects of 360-degree assessment systems on performance and culture

The management literature has examined the performance effects of 360-degree systems, typically comparing the ratings received by a manager in the initial survey to subsequent ratings (e.g., Avery 2000; Smither, London, and Reilly 2005). Literature reviews have concluded that

research has found only modest ratings improvements over time and not among all those appraised (Smither, London, and Reilly 2005; Atwater, Brett, and Charles 2007).[3] A challenge with this research is that 360-degree ratings are often subject to biases: subordinates typically provide overly lenient ratings of their bosses, fearing retaliation (Smith and Fortunato 2008), and self-ratings tend to be even more lenient and less reliable (e.g., Smither et al. 1995; Atwater et al. 2007). The broader literature on subjective performance evaluations by supervisors typically finds centrality bias (which takes place when supervisors give subordinates similar ratings rather than distinguishing good from bad performers) and leniency bias (Prendergast 1999; Bol 2011). Finding improvements in overly lenient ratings can be hard, not only due to low variation but also because ratings are capped at the maximum allowed (ceiling effect).

Another reason for only modest improvements in performance following 360-degree evaluations is that such systems might only work for certain types of individuals or organizations. Differences in personality (such as attitudes towards feedback) and in perceived competence can affect the feedback's effectiveness (e.g., Atwater and Brett 2005; Smither, London, and Richmond 2005). Organizational context could also play a role: in a utility company that implemented 360-degree feedback, employees indicated that they tried harder to improve when they saw greater supervisor and organizational support (Hazucha, Hezlett, and Schneider 1993).

While the empirical literature has paid little attention to the use of 360-degree systems to drive change in an organization's culture and to highlight important organizational values for this purpose, multiple scholars, supported by survey evidence, have suggested such a use (London and Smither 1995). London and Beatty (1993, p. 361) state that such a system "can call attention to

[3] Smither, London, and Reilly (2005) provide several explanations for the small documented improvements, including the crudeness of looking at average ratings at time 2 versus time 1 (the individual might have focused on improving performance related to only a few items), and raters anchoring on their initial impressions.

important performance dimensions heretofore neglected by an organization." Bracken and Rose (2011, p. 188) say that "[a]n organization-wide 360 process that is integrated into its culture can be a powerful tool for communicating and instituting change, rapidly touching all members of the organization when new markets, strategies, values, and structures are introduced into the system."

We build on the research on 360-degree systems by examining the motivational effects of an implementation (a) designed to communicate and reinforce organizational goals and values and (b) in a setting with preexisting explicit incentives. We depart from the existing empirical research, which has focused on mid-level managers in traditional hierarchies, whose compensation presumably is largely fixed pay. Furthermore, rather than examining subsequent ratings, as in prior research, we focus on more objective performance metrics aligned with company goals and values.

2.2 Complementing Explicit Incentive Contracts with Subjective Assessments

The accounting and economics literature has long suggested that organizations using explicit rewards tied to financial results could improve goal alignment by adding subjective performance assessments (Baker et al. 1988; Ittner and Larcker 1998; Prendergast 1999). Financial measures have several limitations: they are transaction-oriented, backwards-looking, and unable to capture the value of many intangible assets affecting future outcomes (Kaplan and Norton 1992; Merchant and Sandino 2009). Even absent an explicit link between subjective assessments and monetary rewards, these assessments are likely to give rise to implicit incentives, since employees will likely anticipate some linkage between the assessments and career-related managerial decisions such as job assignment, salary revision, promotion, and firing (Gibbons and Murphy 1992; Prendergast 1999; Campbell 2008). To the extent that a 360-degree system introduces implicit incentives, it could help align employees with both short-term and longer-term organizational goals, especially employees seeking a career within the organization. Additionally, the multi-source nature of the

360-degree system introduces multi-directional monitoring, which further strengthens the implicit incentive contract (Prendergast 1999; Loughry and Tosi 2008).

2.3 Effects of 360-degree Values-based Assessments in a Setting with Explicit Incentives

Although a values-based 360-degree assessment system could reinforce incentives to exert effort not only towards financial results, but also toward longer-term objectives, employees evaluated on multiple measures may not prioritize the full set required to drive organizational goals. Individuals tend to allocate more effort toward performance dimensions that are more clearly measured, that yield greater results in the short run, and that are associated with monetary incentives (Baker, Jensen, and Murphy 1988; Holmstrom and Milgrom 1991). Faced with seemingly incompatible or difficult goals, they may focus only on the activities they already know how to perform (especially absent appropriate training and supporting systems for new goals).

In sum, the effect of a values-based 360-degree system in a setting with preexisting high-powered incentives is unclear ex-ante. We examine the effects on proxies for performance dimensions corresponding to the values communicated by the 360-degree surveys. We distinguish effects on values closely associated with the financial results rewarded by high-powered explicit incentives from effects on values associated with long-term goals unrelated to monetary rewards.

3. RESEARCH SETTING AND FIELD EXPERIMENT

Our research site, given the fictional name MoPhI, is a mobile phone retail chain in one of India's main cities. A typical store has a manager,[4] a cashier, and a staff of promoters representing various brands (e.g., Samsung, Nokia, and Vodafone) whose products and/or services are offered. To emulate the entrepreneurial spirit, sense of ownership, and incentives of local mom-and-pop stores—the chain's main competitors—the managers and cashiers are compensated mostly with

[4] All store managers were male and typically in their 20s.

sales commissions. The promoters are paid by the brands they represent; they are not MoPhI employees, though they can sometimes participate in some of its sales incentive plans.

MoPhI seeks to differentiate itself from mom-and-pop stores by (a) offering wider selection, (b) bundling products (through promotions) to enhance customers' perceptions of value, and (c) offering custom solutions and trustworthy service. The managing director communicates this value proposition to the store staff through personal visits to the stores, weekly meetings at headquarters, and communications via email and the company's information system. She also strongly enforces the focus on trustworthy service by penalizing—sometimes firing—employees for unethical behavior such as theft and misleading customers.

While a reliance on the managing director's personal interactions with store teams had helped foster a strong company culture, this was clearly incompatible with MoPhI's ambitions for wide expansion. Furthermore, the managing director sought to encourage store teams, particularly managers, to focus not only on short-term financial performance (already strongly incentivized by the compensation structure), but also on longer-term behaviors needed to build a consistent and profitable brand; for example, building long-term customer relations, providing feedback to team members, and not lowering prices just to make a sale. The managing director therefore, in consultation with members of our research team, decided to implement a values-based 360-degree assessment system, centered on the store manager.

3.1 Store Managers and the Values-based 360-degree System

MoPhI store managers, like those in many retail chains (Arnold, Palmatier, Grewal and Sharma 2009), play many roles and are considered key drivers of store success: they lead their store teams, are responsible for forming good, long-term customer relationships, and communicate inventory

and staffing needs to headquarters. They are accountable for the ongoing success of their stores and are expected to model behaviors conducive to the chain's success.

A 360-degree system makes it possible to measure performance with respect to all these roles. Since store staff are best placed to observe their manager in day-to-day operations, their assessments could arguably make for a better performance measurement system than one limited to assessments by the manager's supervisor. As London and Beatty (1993, p. 360) note, "Subordinates … may have more complete and accurate information about many leadership behaviors than supervisors have."

By implementing a values-based 360-degree assessment system, MoPhI's management sought not only to formalize core values and their associated behaviors, but also to gain a complete picture of how each store manager was living those values, so as to provide feedback and coaching. The 360-degree surveys were designed to collect information about each store manager from the store's cashier and brand promoters, the manager's supervisor, the customers, and the manager himself.

The system was intended to be developmental and was presented to employees as a tool to help them grow and increase their chances of promotion (thus, it provided implicit incentives). For the brand promoters, a "promotion" meant being hired as a cashier or store manager. Cashiers were often promoted to store manager and while store managers were unlikely to become district managers,[5] they could be promoted to "store manager coach," to assist nearby stores.

The hope was that with management (a) formally communicating core values, (b) periodically asking store teams and supervisors to consider those values while assessing their store manager's behavior (or, for the store manager, his own), and (c) providing feedback to managers, the teams

[5] District managers are the management layer between the stores and headquarters; most store managers had too little education to deal with the demands of the district manager role. However, the managing director did express a desire to promote store managers to the district manager position under the new system if they performed well on the more traditional performance measures in the company and had a good fit with the core values of the company.

would internalize the core values. Alignment of store teams with core values—that is, the creation of shared beliefs—would increase their identification with the organization. This, in turn, would enable (a) more delegation and less direct monitoring by the managing director (for instance, if store teams could be trusted to put the company's long-term success ahead of short-term financial interests) and (b) greater work satisfaction and motivation for store teams (Van den Steen 2010).

3.2 Inputs to the System

The main surveys associated with the 360-degree system asked each respondent (the store manager, store members, and the supervisor) to assess the manager on various behaviors, organized by four core values: (a) "We gain control of our own career by working hard every day and reaching out for support"; (b) "We give more value"; (c) "We are honest and ethical"; and (d) "We are caring and respectful." Respondents were also asked where the store manager was doing a good job and where he could do better. The survey instrument is shown in Appendix 1.

One customer per store was randomly selected each day for a telephone survey. They were asked about aspects of the service they received (such as the staff's knowledge and politeness) and how likely they were (on a scale of 0 to 10) to recommend MoPhI to a friend. From this question, a store's *Net Promoter Score* can be calculated as the percentage of "promoters" (respondents giving a 9 or 10) minus the percentage of "detractors" (respondents giving a score of 0 through 6).[6] The *Net Promoter Score* is a popular metric for customer experience.

3.3 The Field Experiment

Since the managing director was interested in understanding the effectiveness of a values-based 360-degree assessment system (developed and implemented via extensive consultation with the research team), she readily agreed to introduce it as a field experiment, initially in

[6] See www.netpromoter.com.

approximately half the stores.[7] While she was enthusiastic, it was certainly not a forgone conclusion that the system would drive behaviors aligned with long-term organizational goals and values. The system wouldn't work, for example, if employees ignored it because it was not linked to high-powered explicit incentives, if they did not understand or embrace the long-term goals, if biased responses made the feedback to store managers less meaningful, or if managers expected their subordinates to inflate their ratings and therefore saw no reason to change their own behavior.

Store selection was randomized. However, we grouped stores in "blocks" if they were close enough to each other that contamination effects would be a concern if some participated and others did not. If any store in a block was randomly selected, all stores in the block were then selected. Participants in the selected stores were advised that MoPhI was piloting the system only in certain stores and asked not to discuss it outside their store team.

The system was launched with an inspiring presentation and interactive session led by the managing director. The session was held twice, at the end of March 2015 and at the end of April 2015, to include all the selected stores and to help some stores that had difficulties accessing the 360-degree surveys online. The presentation had two parts. The first introduced store teams both to the company's competitive strategy and to its newly formalized vision statement and core values (included in Appendix 2), emphasizing the importance of the role of the store managers and store teams to accomplish these values and goals. During the second part, the managing director explained that a 360-degree system was to be implemented at the stores in attendance in order to gain a comprehensive understanding of the support the store manager provided and his commitment to the core values. Attendees were advised that the aggregated survey responses

[7] This opportunity to collaborate with MoPhI arose due to a preexisting relationship between members of the research team and the managing director. They had met her at a retailing conference and subsequently built up a strong relationship.

(confidentiality of individual responses was assured) would be used to provide feedback and coaching to store managers. Attendees were then asked to complete the performance appraisal survey (Appendix 1) in a computer lab that were set up for this purpose.[8]

By the end of the second session, 20 of MoPhI's stores had participated in the launch of the 360-degree system.[9] The feedback sessions were held about three months after the second session. By this time, a couple of treated store managers had moved to control stores (we speak to this further in footnote 14) and a couple of feedback sessions that should have taken place did not, reducing our available sample for the period post the feedback sessions. Further complicating matters, there was some variation in how these feedback sessions were conducted (e.g. whether or not the managing director was present). Hence, we end our post-period right before the feedback sessions.[10] This unintendedly allowed us to examine the motivational effects of the 360-degree system implementation, independent of any feedback effects.

4. RESEARCH DESIGN AND EMPIRICAL TESTS

While prior research has studied the performance effects of 360-degree systems by comparing initial ratings with subsequent ratings, we do not study a change in ratings because 360-degree surveys were completed only once during our sample period. However, studying a change in

[8] The number of items on survey instruments for upward feedback (in which subordinates provide feedback to their superior) and 360-degree feedback systems can vary widely. For instance, the survey in Hazucha, Hezlett, and Schneider's (1993) study had 122 items, while the survey in Walker and Smither's (1999) study had 29 items.

[9] In our empirical analyses we drop all weeks between the first and second session.

[10] The absence of feedback in the post-period that we study may have affected the results that we document, but research suggests this is not a significant issue. Smither, London, Vasilopoulos, Reilly, Millsap and Salvemini (1995) find that managers who received individualized feedback in an upward feedback system were no more likely to improve their performance than managers who did not. (In their setting, managers with fewer than three subordinates received only an aggregate organizational report to protect the subordinates' anonymity.) Using the same managers as in Smither et al. (1995) but studying them longer, Reilly, Smither, and Vasilopoulos (1996) also find that the performance improvements they identify over four administrations of the feedback system (mostly concentrated early in the system) are unrelated to the number of times managers actually received individualized feedback (which varied from zero to three). They conclude: "Our results suggest that the continued administration of an upward feedback program can result in sustained change over a fairly long period of time and actually receiving feedback may be less important than the exposure to the valued behaviors" (p. 599).

ratings would likely have been relatively uninformative since the surveys completed by the store managers and the store team members (cashier and promoters) were extremely favorable; store managers rated themselves an average of 4.8 across all survey questions, while team members gave them an average of 4.4. Their direct supervisors, however, gave them an average of 3.8.

Our interpretation of these highly favorable ratings, partly informed by conversations with the managing director, is that it was due to several factors, including store managers wanting to paint themselves in the most favorable light; a lack of familiarity with surveys in India; the hierarchical nature of Indian society, leading team members to be hesitant to reveal anything negative about their manager; and possible influence activities whereby the store managers asked their team members to give them high ratings. Often, when 360-degree feedback is used for developmental purposes, the feedback is provided only to the individual, not to his or her manager. In our setting, however, the small size of the organization and the lack of familiarity with such systems necessitated the managing director's involvement.

While we cannot study a change in ratings, we test the effects of the intervention on performance measures that were readily available (or could be easily constructed) from the organization's existing information system.[11] Our approach allows us to examine the effects of the intervention on measures of performance that are (a) related to the core values introduced with the 360-degree system, (b) more objectively measured than those used in prior studies, and (c) less subject to leniency and centrality bias. Furthermore, the availability of a treatment and a control groups allows us to better assess causal relations relative to prior studies relying predominantly on earlier and later performance assessments after the system had been introduced. Our empirical

[11] All of the measures we examine pertain to information that had historically been recorded with the exception of the customer surveys, which began shortly before the introduction of the 360-degree feedback system and were the result of our collaboration with the company.

analyses explore the performance effects of the implementation of a values-based 360-degree system, which included the formalization and communication of the company's vision and values.[12] We describe our research design in the next section.

4.1 Research Design

We examine how treated stores performed vis-à-vis control stores on three dimensions capturing behaviors related to the core values assessed through the 360-degree system: (1) performance on metrics included in the existing explicit incentive system, associated with the first core value of working hard; (2) performance on productive behaviors not related to explicit incentives, associated with the second core value of providing value to customers; and (3) counterproductive behaviors, associated with the third core value of being honest and ethical. (Appendix 2 lists the measures we use to capture these dimensions of performance and their corresponding core values). No suitable objective measures were available to evaluate performance on the fourth core value, "*We are caring and respectful.*" Therefore, we excluded this dimension of performance from our analyses. All measures were discussed with the managing director, who approved them as appropriate proxies for the underlying aspects of performance she intended to emphasize with the 360-degree system.

To capture performance metrics already included in the preexisting monetary incentive system and in line with the core value, "*We gain control of our own career by working hard every day and reaching out for support,"* we define *LogSales* and *LogGrossProfit* as the natural logarithms of store-level sales and of gross profit, respectively. Metrics aligned with the core value "*We give more value*" and not explicitly rewarded with monetary incentives include the net promoter score (*NPS*) and the percentage of invoices including promotion-related bundles (*BundleInvoice*), both

[12] In our setting, we cannot disentangle the effects of communicating the company's vision and values from the effect of introducing them through a 360-degree system. We leave this for future research.

captured at the store-week level. With respect to the core value "*We are honest and ethical*," the variable *AbnormalReturns* measures the abnormal incidence of customer returns in the first week of every month relative to other weeks in the month (a proxy for gaming behaviors, whereby sales associates seeking to increase their monthly commissions sell additional units at the end of the month that are then returned at the beginning of the next month).[13]

Our sample includes weekly observations for 32 stores (16 each in the treatment and control groups) spanning 22 weeks, of which 11 are prior to the initial launch (pre-period) and 11 are subsequent to the re-launch (post-period).[14, 15] We drop any post-period observations subsequent to the departure of a treated store manager if there was no replacement and any post-period observations in which a control store is contaminated by a treated store manager being reassigned to it. Our final sample includes 693 store-week observations. We have fewer observations for our NPS measure, since the customer surveys began only three weeks before the 360-degree system was first launched. Also, the metric related to gaming can only be estimated on a monthly basis.

We examine the effect of the intervention using the following difference-in-differences statistical model:

$$Performance_{i,t} = \alpha + \beta_1 Post_t + \beta_2 Post_t * Treatment_i + \beta_3 StoreManagerChange_{i,t} + \beta_4 SalesDays_{i,t} + \beta_n(Store\ Fixed\ Effects) + \varepsilon, \quad (1)$$

[13] The company also performed periodic random audits of the correspondence between (a) inventory and cash represented in the local ledgers and (b) their physical presence in the store. Store managers were directly responsible for any unfavorable deviation and were penalized with pay deductions equivalent to the value of the missing asset. Stores were audited on average once a month. Due to low power to detect changes in this variable and because a number of stores were audited only in either the pre- or post-period, we did not include this measure in our analyses.

[14] Our initial sample included 39 stores. Of these, seven were dropped. One was eliminated because its manager did not complete a 360-degree survey (presumably because he was temporarily reassigned to a control store in April 2015), making it uncertain whether the store should be classified as part of the treatment sample. Another store, which opened immediately before the sample period and closed immediately after it, was dropped due to several problems that made this store different from the rest of the sample. In the case of two stores, the store team participated only in the core values session and not in the 360-degree survey. One control store was contaminated by the reallocation of a treated store manager immediately after the intervention. One store was included in the treatment group, but the store manager did not fill out the survey. One store team identified the cashier as the (acting) store manager, but the cashier himself did not self-identify as store manager in his survey.

[15] Our pre-period includes the 11 weeks preceding the first launch to keep correspondence between the length of the pre-period and the length of the post-period, in line with Charness and Gneezy (2009).

where the dependent variable *Performance* is substituted by each of the dependent variables described above; the indicator variable *Post* equals 1 if week (month) *t* is after the implementation of the 360-degree system, and 0 otherwise; the indicator variable *Treatment* equals 1 if the system is implemented in store *i*, and 0 otherwise;[16] the indicator variable *StoreManagerChange* equals 1 if the manager of store *i* was different in week (month) *t* than at the time of the intervention, and 0 otherwise; and *SalesDay*s is the number of days in the week or month (whichever corresponds to the dependent variable's unit of analysis) during which that store is open.[17] We also include store fixed effects. All variables included in our statistical analyses are described in Appendix 3.

Table 1 summarizes descriptive statistics for the variables of interest. Sales and profitability measures exhibit significant variation across store-weeks. Despite some instances of a perfect NPS score (100%), NPS is at or below 33% for half of the sample.[18] The percentage of promotion-based bundles (*BundleInvoice*) is below 34% for three-quarters of the sample, suggesting opportunities to offer greater customer value. Dysfunctional behaviors seem pervasive, further justifying the need for the intervention. *AbnormalReturns* is equal to or greater than 50% for a quarter of the observations. Stores operate, on average, 6.5 days per week and 28.2 days per month.

----- Insert Table 1 here -----

4.2 Results of statistical tests

4.2.1 Univariate analyses

Table 2 reports the Pearson correlation coefficients calculated across all of our variables of interest. As expected, there is a very high correlation between our two financial measures of

[16] The main effect of *Treatment* does not appear in Equation (1) since it is absorbed by the store fixed effects.
[17] Variation in the number of sales days for the same time period (week or month) across stores is explained by differences in local market dynamics (busier and wealthier neighborhoods might expect stores to be open on weekends), or demographics (e.g., stores in Muslim neighborhoods are likely to close on Muslim holidays).
[18] In the analysis of NPS, we eliminate two stores for which we had only post-period observations. We also calculated alternative versions of the NPS variable in line with Casas-Arce, Lourenço, and Martínez-Jerez (2017) and estimated all statistical models with these alternative measures, obtaining equivalent results.

performance related to explicit incentives (*LogSales* and *LogGrossProfit*). These metrics are also positively correlated with cross-selling (*BundleInvoice*). On the contrary, sales, profitability, and customer satisfaction (proxied by NPS) are negatively correlated with *AbnormalReturns*, suggesting that financially healthy stores are less likely to engage in dysfunctional behaviors such as gaming. Our overall assessment of the correlations among our predictors is that the risk of collinear relations in the definition of our statistical model is not material.[19]

----- Insert Table 2 here -----

We report the results of our univariate tests in Table 3. We perform paired t-tests comparing mean performance of the pre- and post-periods (reported in the columns of Table 3) and of the treatment and control group stores (reported in the rows). The bottom right corner provides a univariate difference-in-differences analysis (comparing the pre-post differences in values of the dependent variables of the treatment versus the control groups). While we refrain from drawing conclusions on the effects of the intervention, the univariate results do provide relevant insights.

First, despite our random assignment of stores to the treatment and control conditions, we notice greater pre-intervention sales and profits for the treatment group. Second, we shed light on several interesting trends with respect to our variables of interest. Univariate tests with respect to changes in dimensions of performance that are not rewarded in the preexisting incentive system yield contrasting results: one of our customer-related variables (*NPS*) improves significantly in the post-period, while the other (*BundleInvoice*) deteriorates. However, we cannot conclude that these changes are necessarily a consequence of the 360-degree system, because we observe similar trends in both treated and control stores. Regarding the intervention's effects on counterproductive

[19] We corroborate our conclusion by running several tests of collinearity in the estimation of the coefficients in our model. For example, VIFs are all below 2.

behaviors, we find a decrease in *AbnormalReturns* for both treated and control stores—possibly driven by tighter monitoring in the post-period—but no significant differences between them.

----- Insert Table 3 here -----

4.2.2 Multivariate analyses

Table 4 summarizes our multivariate analyses. We perform separate estimations of Model (1) for each of our dependent variables of interest, using OLS regressions. To account for the fact that we observe repeated measures of performance for each store in our sample, all our estimations include store fixed effects (which serve as a control for the store's pre-intervention values of the outcome variables) and standard errors clustered by store.

Since Model (1) is specified using a difference-in-differences approach, our primary coefficient of interest is β_2, which is associated with the interaction term (*Post*Treatment*). β_2 is positive and significant with respect to *LogSales* (β_2=0.211, p<0.10) and *LogGrossProfit* (β_2=0.260, p<0.10).[20, 21] In terms of economic significance, this suggests that the intervention was associated with 23% higher sales and 30% higher gross profits in treatment stores, relative to control stores.[22] Thus, we conclude that the introduction of the 360-degree system motivated store teams to try to increase financial performance above and beyond the effort induced by the existing incentive plan. This result is consistent with the 360-degree system having a positive effect on

[20] All our p-values are calculated based on two-tailed statistical significance.

[21] We note that the main effect of *Post* is negative and statistically significant, suggesting a general decline in sales and gross profit after April 2015. The placebo tests described in footnote 25 also document a negative main effect of *Post* on *LogSales*, suggesting that a similar trend might be common to the firm's annual cycle. This suggests that even a successful intervention could only have slowed down the decline in performance in our post-period and that actual increases in sales would be unlikely.

[22] This economic significance may be an upper bound because, as we explain in footnote 25, we found a positive but insignificant interaction for *Post*Treatment* in a placebo test using prior-year data. We cannot completely rule out that the economic magnitude we capture is a result both of seasonal trends that may, for some reason, differentially affect treated stores (perhaps because they sell more than the control stores) and the intervention's effects.

measures of performance previously rewarded with monetary incentives by increasing the salience of "*Working hard every day*" and/or introducing implicit incentives. [23, 24]

Surprisingly, we find no significant differences between treated and control stores in the post-implementation period with respect to measures of performance not included in the preexisting incentive system (*NPS* and *BundleInvoice*) or with respect to our gaming behavior indicator (*AbnormalReturns*). A main objective for the 360-degree system was to shift the store managers' allocation of effort toward organizational values focused on the firm's long-term sustainability. Our results, however, do not provide evidence of a significant change in that direction. Our findings may be interpreted in line with prior research on multitasking environments, whereby employees might react to the introduction of multiple measures of performance by focusing on those activities that are already well understood, more clearly and precisely measured, more sensitive to effort, and more directly linked to monetary rewards (Baker, Jensen, and Murphy 1988; Holmstrom and Milgrom 1991; Ittner and Larcker 1998; Dikolli and Sedatole 2007; Tayler 2010).

[23] Untabulated analyses show that our results are robust to the inclusion in the sample of the store (originally excluded from the dataset) that opened immediately before our sample period and closed immediately after it. Additional untabulated analyses show results that are comparable in size but their significance is reduced when including time fixed effects, most likely due to lack of power: the coefficient on *Post*Treatment* for the sales regressions becomes 0.20 (t=1.63) when including week dummies and 0.20 (t=1.65) when including month dummies and, for the gross profit regressions, becomes 0.23 (t=1.66) when including week dummies and 0.23 (t=1.67) when including month dummies.

[24] To test the robustness of our results, we perform several placebo tests. Because of the tight correlation between sales and gross profits, we focus our placebo tests on the estimation of the model predicting *LogSales*. First, we re-randomize the stores in our sample so that 50% of the stores originally in the treatment (control) group are artificially reclassified into the control (treatment) group. As expected, estimating Model (1) using the re-randomized sample yields a nonsignificant coefficient associated with the interaction term (t = 0.60; p = 0.553). We also estimate a simplified version of Model (1)—we exclude store manager changes for 2014 because we did not construct that variable—using data for the same sample of stores and for the same weeks in the year prior to the intervention (2014), assuming that the intervention occurred in 2014. The estimation yields a positive but non-significant coefficient associated with the interaction term (t = 1.33; p = 1.19). Running the same simplified regression on our sample (2015 data) continues to yield a significant coefficient (t = 1.73; p = 0.09). We note that the regression estimated with 2014 data yields a negative and significant coefficient associated with the main effect of *Post*, suggesting that sales might be subject to seasonal trends. Finally, we estimate the full Model (1), including lagged sales (i.e., sales corresponding to the same week in 2014) as a predictor. This specification, which includes store and market characteristics (time-invariant in our main tests) and excludes store fixed effects, yields a positive and significant coefficient for the interaction term (t = 2.24; p = 0.03) and we find results equivalent to those of our main estimation.

Qualitative information obtained from interviews provides additional insights, described in Section 5, into other possible causes of the 360-degree system having no discernible impact on the nonfinancial measures or dysfunctional behaviors. Additionally, given our small sample size, some of the results may not have been captured by our estimations due to low statistical power. We discuss potential power issues and their impact on our results in Appendix 6.

Taken together, our empirical tests depict a somewhat counterintuitive result. The intent of a values-based 360-degree assessment system was to increase effort on dimensions supporting long-term organizational success and the intervention did indeed incentivize greater effort, but not in a uniform way across all dimensions of performance. At the same time, we note how the favorable effect documented with respect to the measures of financial performance did not come *at the expense* of the performance aspects related to long-term sustainability, as we do not find any significant deterioration in the productive measures capturing customer service that may lead to future business and no increase in gaming behaviors.

----- Insert Table 4 here -----

5. QUALITATIVE ANALYSES AND SUPPLEMENTARY ANALYSES

In this section, we discuss the follow-up interviews and supplementary analyses we conducted to better understand our results.

5.1 Qualitative analyses

In late May and early June 2015 (that is, a little over a month after the system launch and the survey), we requested an assistant in India (identified by the managing director) to conduct follow-up interviews on our behalf to gauge the sentiment of MoPhI's employees with respect to the 360-degree system. Fourteen store staff members (two store managers, three cashiers, and nine

promoters) from five stores randomly selected from the treated stores were interviewed using a structured interview questionnaire (see Appendix 4).[25]

Descriptive evidence gleaned from interviews indicates that MoPhI employees viewed both the session launching the 360-degree system and the completion of the related surveys positively. They understood that the system involved obtaining feedback about the store manager and many stated that the purpose was to provide greater transparency. Some respondents mentioned motivating the store manager and/or helping the store manager to grow/improve as purposes of the system. Expected beneficial impacts mentioned by respondents included motivating employees, improving the store environment and teamwork, facilitating support from senior management, and developing employees. In our conversation with the managing director, she highlighted that the brand promoters were especially interested in the first core value, related to gaining control of one's own career, and how they could work towards being promoted to cashier or store manager.[26]

Despite the positive perceptions of the system, nearly all interviewees said that they could not recall the core values.[27] The exceptions were: an interviewee who recalled one core value correctly, and not the others; an interviewee who recalled one core value correctly and gave a second but incorrect core value, though it did include elements of two actual ones; and an

[25] One of these treatment stores was excluded from our empirical analysis because the store opened immediately before and closed immediately after our sample period (as explained in Section 4.1). Furthermore, for this particular store, our records indicate that a promoter was in the process of becoming a store manager (we include that employee in the count of store managers above) right around the time of the system launch, which appeared to generate some confusion as to who the survey respondents should be considering as the store manager. Nonetheless, because of our small number of interviews, our findings in this section include the interviews with individuals from that store.

[26] She explained that a brand promoter who achieved his or her target was typically given a higher target set by the brands the next month and had limited opportunities for career advancement.

[27] Note that the initial sessions in which the 360-degree system was introduced were the only times when store employees were exposed to the new core values. To prevent contamination of the control stores, the core values were not disseminated after the launches. This feature of the rollout, which was necessary for the purposes of the field experiment, likely contributed to the employees' limited recall. Furthermore, the extensiveness of the 360-degree survey may have overwhelmed them. Our findings may also be somewhat impacted by the sample of interviewees: a few had not attended the launch session, but had completed the survey after the system was explained to them later, and only two store managers were interviewed.

interviewee who listed four core values, none of which were correct but which were somewhat related to the core value of working hard. When we spoke with the managing director immediately after the launch of the 360-degree system, she noted that the employees were very interested in how to create long-term relationships with their customers (related to the core value "*We give more value*"): "They are asking, 'How do we live it daily? We have too many customers that walk into our stores. So how do we build a relationship with our customers? How would you do it?' . . . One guy asked me, 'How is it done in the US in Starbucks? . . . We don't know. Is there any study we can read? Is there anything we can learn from other folks?'"

Surprisingly, more than a third of the interview respondents indicated that they would not recommend working at MoPhI to their family or friends. Twenty-nine percent complained about salaries not being paid on time, 21% complained about lack of support (though usually highlighting that this did not extend to the managing director, who was considered supportive), and one mentioned that inventory wasn't always provided as needed. We therefore conjecture that store managers, feeling that the organization demonstrated no clear commitment to either the core values or its employees, might have considered management's request that they take on additional responsibilities related to the core values for the sake of MoPhI's long-term success to be unfair. Furthermore, store managers may have lacked the tools and knowledge necessary to drive improvements in long-term relationships with customers, even if they wanted to.

5.2 Supplementary analyses

To further explore the role of employees' concerns about their inability to execute some of the core values (especially those related to driving customer value) and about their lack of support, we describe the results of supplemental analyses examining two potential moderators of the effect of

the assessment system: store manager tenure (to proxy for ability) and availability of inventory (to proxy for support).[28, 29]

As MoPhI has grown, the managing director has had less opportunity to directly influence and monitor the behaviors of store employees. While some store managers have been with MoPhI a long time and have been exposed to its core values by working side-by-side with the managing director, many others, hired more recently, have not had such exposure.

We explore whether the tenure of the store manager moderates the effects of introducing a values-based 360-degree assessment system. On the one hand, store managers who have been with the organization for a longer time may already have internalized the core values in their daily activities. Therefore, compared to more recent hires, they may exhibit a muted reaction to the introduction of a system geared towards raising awareness about those values. On the other hand, store managers with longer tenure might have greater expertise in executing work according to those values, expertise which they can transfer to their sales teams. For instance, Griffith and Neely (2009), studying a large distribution firm that was introducing a Balanced Scorecard with a large number of measures, find that only experienced managers could interpret the measures and improve performance. Similarly, in a study of a convenience store chain, Campbell et al. (2015) find that when faced with additional tasks, low-skilled employees were unable to perform. In the specific context of a 360-degree system, Atwater and Brett (2005) find that the system was associated with better results when the leaders considered themselves to be more competent.

[28] The main effects of these characteristics, which are time-invariant for the purposes of this study, were subsumed by the inclusion of store fixed effects in the estimations of all our main statistical models. These additional analyses, too, include store fixed effects, but also interaction terms between these characteristics and the variables representative of the intervention (namely *Post* and *Treatment*).

[29] We also considered geographic distance in miles between the local store and the headquarters as a potential moderator associated with familiarity with company values. Our statistical analyses (untabulated) yielded no significant coefficient for our variables of interest.

As a result of MoPhI's introduction of the values-based 360-degree system, teams reporting to store managers with longer tenure may have become more sensitized to organizational values. Tenured store managers are also likely to have greater interest in developing their careers at MoPhI than newcomers, which might increase their sensitivity to the implicit incentives associated with the system. To examine whether store manager tenure moderates the relation between the introduction of a 360-degree system and performance, we estimate the following model:

$$Performance_{i,t} = \alpha + \beta_1 Post_t + \beta_2 Post_t * Treatment_i + \beta_3 Post_t * Tenure_i + \beta_4 Post_t * Treatment_i * Tenure_i + \beta_5 StoreManagerChange_{i,t} + \beta_6 SalesDays_{i,t} + \beta_n(Store\ Fixed\ Effects) + \varepsilon, \quad (2)$$

where *Tenure* is the number of months of employment recorded for the store manager, computed at the time of the introduction of the 360-degree system. A positive and significant estimation of β_4 would indicate that the introduction of a values-based 360-degree assessment system does more to improve performance the longer the store manager has worked with the organization.

The estimation results are reported in Table 5. While we find no evidence of a moderating effect of tenure with respect to financial metrics already rewarded, we find that tenure moderates the performance effects with respect to *NPS* (β_4=0.002, $p<0.05$).[30] This suggests that for every year a store manager was with the company, the effect of the 360-degree system was associated with a shift of 2.4 percentage points (0.002*12) from % of detractors to % of promoters.

----- Insert Table 5 here -----

Next, we analyze the moderating effect of the availability of inventory to examine the role that support could have played on the effects of the 360-degree system. Margins for mobile phone sales are generally small and procurement of inventory requires significant investment in working capital; the devices are largely pre-paid by the retailer and risk obsolescence and damage if kept

[30] Our results are robust to restricting the sample to stores whose store managers did not change during our sample period.

in-store too long. Consequently, MoPhI's management tries to minimize the stock on hand, but a consequence is that stores may not always have what the customer wants. Devices can be transferred from one store to another relatively easily, but lack of availability and waiting time might nevertheless reduce customer satisfaction, the likelihood of a sale, and the possibility of bundling—that is, the performance outcomes associated with the first two core values of working hard and giving more value to customers. We use the following statistical model to explore whether inventory availability moderates the performance effects of the 360-degree system:

$$Performance_{i,t} = \alpha + \beta_1 Post_t + \beta_2 Post_t * Treatment_i + \beta_3 Post_t * HighestDSI_i + \beta_4 Post_t * Treatment_i * HighestDSI_i + \beta_5 StoreManagerChange_{i,t} + \beta_6 SalesDays_{i,t} + \beta_n(Store\ Fixed\ Effects) + \varepsilon, \quad (3)$$

where the indicator variable *HighestDSI* equals 1 if the store is in the highest quartile of the distribution based on days sales in inventory (DSI) in the pre-period, and 0 otherwise. In line with practice, we use DSI as a proxy for the availability of inventory on hand. DSI is calculated as the ratio of average inventory balance to the average cost of goods sold per day; it represents the number of days required to convert the stock on hand into sales in a given period. A higher DSI indicates greater availability of inventory on hand. We measure average DSI for each store along the weeks included in the pre-period.[31] We then partition the distribution of pre-period average DSI into four quartiles. Our assumption is that inventory availability might not influence store performance in a continuous manner; rather, there might be discontinuities in the relation. Since the company placed more inventory in certain stores, to use them as hubs, we assumed that those stores would be better equipped than others to improve their performance in response to the system.

[31] Our choice is informed by the possibility that changes in sales performance subsequent to the implementation of the 360-degree system might impact the DSI measure even in the absence of changes in inventory management policy.

We identify those stores as those with the highest DSI (at the 75th percentile of the distribution).[32] In estimating Model (3), a positive and significant β_4 would indicate that stores with the most inventory improved performance to a greater extent than stores with less inventory after the introduction of a values-based 360-degree assessment system.

Table 6 reports the estimation results. Our estimations related to the interaction between *Post*, *Treatment*, and *HighestDSI* indicate that stores with the most inventory perform better with respect to bundling products (β_4=0.131, p<0.10), thus providing greater value to customers through promotions and cross-selling. This suggests that, among stores with the highest DSI, the effect of the 360-degree system was associated with a 13% increase in the number of invoices including bundles in treatment stores compared to control stores. We found no effect on other treated stores.

----- Insert Table 6 here -----

The results of this study were illuminating to the managing director, who then took steps to further promote and clarify the company's values and to improve employee support, including paying salaries on time and increasing support to store managers through coaching and systems.

6. LESSONS LEARNED

Despite the promise of field experiments as a way to achieve both randomization and realism and as a complement to laboratory studies and archival research, their use in accounting has been limited (Floyd and List 2016).[33] Field experiments are more common in economics and, drawing on his experiences conducting such studies, List (2011) offers 14 tips for executing them, presented in Appendix 5. His tips resonate with us. The commitment of the managing director to the 360-degree system, her investment in it, and her interest in understanding its effects definitely

[32] In untabulated analyses, we adopted a discontinuity at the median value of the distribution of DSI and found no significant moderating effects.
[33] Recent examples include Casas-Arce, Lourenco, and Martinez-Jerez (2017) and Li and Sandino (2018).

contributed to our ability to work with her to implement it as a field experiment and to obtain data and access to personnel. Although, the organization's relatively small scale meant that we had low power to detect effects in some of our dependent variables (see Appendix 6), we felt that this was outweighed by the potential upsides: positive change in the organization and an opportunity to document it, a stronger relationship with the organization that may pave the way for future field experiments (especially once the organization reaches a larger scale), and the opportunity to learn more about how to conduct a field experiment.

In the spirit of disseminating best practices and lessons learned in conducting field experiments, we build on List's (2011) tips and offer the following additional tips stemming from our own experience with this experiment, which may be especially applicable to experiments within a single organization.

A. Be mindful of the possible trade-off between the strength of the treatment and the risk of contamination.

Since all of the stores involved in our field experiment are in one city in India, we were concerned about contamination of the control stores. One measure we took to mitigate this risk, as described in Section 5, was to prevent the dissemination of core values after the participants were exposed to them and completed the survey during the launch session (they were neither able to take a printed copy of the core values or to view the survey after it had been completed). The treatment stores were also advised that the system was being piloted and were asked not to discuss it beyond their store. While these steps likely helped to reduce contamination, they also weakened our treatment. Were it not for concerns of contamination, the organization could have disseminated the core values and reinforced them and the 360-degree system at the weekly meetings at headquarters, through the information system, through store displays, and in other ways. Not all field experiments will involve such a trade-off, but we urge researchers to (a) consider whether

contamination can occur and, if so, through what avenues (such concerns are likely to become even more pronounced with the increasing use of social media) and (b) balance this as well as possible with the strength of the experimental treatment.

B. When the underlying mechanism(s) are hard to identify or tease out empirically, consider gaining insights through interviews.

Results of field experiments sometimes raise follow-up questions. Interviews with participants in the treatment group can be helpful in revealing possible explanations for the empirical results, which could spur additional analyses (such as those using moderating factors) and/or future research designed to test propositions that arise from the interviews. Be sure to consider which participants make the most sense to interview. For our experiment, we sought a mix of perspectives by interviewing store managers, cashiers, and promoters. On reflection, however, since it was ultimately store managers who were directly treated and would drive store results, we could have concentrated our efforts on them. We only interviewed two, both of whom had little recall of the core values (it is, however, difficult to extrapolate this to the total population of treated managers).

C. Be on the lookout for cross-country cultural differences.

We believe there is much to be gained from studying the effects of management control systems both within and outside the United States. Since the opportunity to conduct a field experiment in an organizational setting often stems from an existing relationship between a researcher and an organization, the researcher has to be willing to accept that cultural aspects of the country in which the organization operates may play a role in the effects of the experiment. While we don't see a problem with this, we do believe it is worthwhile for researchers to point out factors that they believe may have played a role. For instance, the hierarchical nature of Indian society combined with a lack of familiarity with surveys—specifically, with 360-degree system surveys—are likely to have contributed to the lenient ratings on the 360-degree surveys. Pilot

testing the surveys also revealed some things that got "lost in translation" into Hindi and which we subsequently corrected, so researchers should be careful to use high-quality translators and to translate back and forth if parts of the experiment are conducted in another language.

D. Give some thought to creating a contract of sorts with the research site.

While we discussed the purpose and benefits of a field experiment with the research site at length, we now believe that a nonbinding contract would have helped avoid some of the pitfalls we encountered. Such a document could explain in detail the design features needed for a field experiment to work properly and what the research site and the researchers need to do to fulfill their ends of the bargain. Both the research site and the researchers should sign off on this document. As we noted earlier, our feedback sessions were not held until a few months after the launch of the 360-degree system. This lag was not intended, but arose due to the considerable time the organization took to compile the survey responses, prepare for the feedback sessions, and so on. Getting the organization to commit to a timeline and plan for the resources involved at each stage would likely have allowed us to examine the effects of the feedback sessions and to limit concerns associated with the system potentially "losing steam" with the employees due to the absence of timely feedback. Furthermore, reassignments of treated store managers to control stores in the post-period forced us to drop certain store-months from our sample. Such issues are more likely to arise in fast-moving, dynamic organizations like the one we studied, but having an upfront detailed agreement could have helped us preempt this type of contingency.

E. Check the covariate balance between the treatment and control group prior to running the experiment, especially in small samples.

Random assignment of the unit of analysis to treatment and control groups is designed to take care of any unobservables and should also result in covariate balance between observable variables. In smaller samples, there is a greater risk that randomization will not completely

accomplish this, as we discovered when comparing differences in pre-period outcome variables between our treatment and control groups. While we don't believe this is a significant concern, having included store fixed effects in our empirical analyses, we would recommend checking the covariate balance at the time of randomization and re-randomizing if necessary.

7. CONCLUSIONS

This study explores the effectiveness of a values-based 360-degree performance assessment system, introduced in tandem with a formalized vision and core values, as a means to improve alignment in an organization already motivating its employees via incentives tied to financial performance. Such a system was designed to motivate employees toward behaviors aligned with longer-term organizational objectives—while maintaining or increasing effort devoted to financial performance—through communicating organizational goals and values, creating implicit incentives, and reinforcing intrinsic motivation. However, such a system would not work if, for instance, employees ignored it because it was not linked to high-powered explicit incentives or if they did not understand or embrace the long-term goals.

We developed and implemented a field experiment in collaboration with a mobile phone retail chain in India; its incentive system already rewarded behaviors related to sales and short-term profitability with substantial monetary awards. The intervention, by shifting employee attention towards behaviors aligned with long-term organizational sustainability, aimed to correct for the distortions the current incentive system had introduced. As part of the experiment, management formalized and communicated a company vision and core values and introduced a 360-degree system in half its stores. The preexisting incentive system was not modified and no additional explicit rewards were introduced during our sample period. We examine the effects of the

introduction of the system, before any feedback was provided, on outcomes that were largely controllable by employees in their day-to-day work.

Results of our statistical analyses show that the intervention yielded somewhat counterintuitive results. On the one hand—and consistent with our expectations based on prior theory—introducing a values-based 360-degree assessment system had a positive effect on effort allocated to dimensions of performance that are more clearly measurable, more subject to employees' control, and associated with explicit monetary incentives and that yield greater short-term results. Despite an unaltered reward system, performance related to sales and profitability improved in the stores that were exposed to the 360-degree system. On the other hand, the system did not fully accomplish its main purpose of aligning employees' goals toward dimensions of long-term organizational performance and sustainability. Insights obtained from employees suggest that low levels of recall of the core values, inability to act on some values, and inadequate support from the organization to achieve the new organizational goals might have contributed to our findings. Additionally, moderation tests show that factors such as store manager tenure (capturing ability) and inventory availability (capturing support) influence the intensity of the outcomes of the intervention.

Our findings contribute to the literature on multi-source performance appraisal systems by highlighting both their benefits as drivers of performance absent additional monetary incentives and potential side effects that might weaken their effectiveness and result in suboptimal allocation of effort on some of the dimensions they promote. We contribute to the study of performance in multitasking conditions in a setting in which financial incentives are associated with only some of the desired behaviors. Finally, we contribute to the practice of management accounting by documenting benefits and pitfalls of introducing a values-based 360-degree system.

Our results are generalizable to companies considering a 360-degree system in a setting with preexisting high-powered incentives (though being focused on a single organization, our study is subject to the usual concerns regarding generalizability). We acknowledge that low-powered tests and potentially imperfect proxies for our core values (the impact of which we discuss in detail on Section 4) may have hampered our ability to find significant results for *some* outcome (though not all) variables related to longer-term organizational goals. However, our proxies allowed us to use more objective measures than the ratings traditionally used to test the effects of performance assessments interventions and to capture potential causal effects (as we could compare the outcomes of the treated group with those of a control group). Additionally, the cultural context of India—its hierarchical society and the lack of familiarity with surveys—may have made the 360-degree system's impact less than desired. Yet we believe there is much to be learned from introducing management control systems in different countries and cultures and identifying the factors that may contribute to or hamper their success.

Future research may address the conditions under which a 360-degree system focused on organizational goals and values will drive improvements in outcomes not linked to financial incentives. For instance, how is an employee's willingness to exert effort towards behaviors that are not rewarded in the short-term affected by his or her intrinsic motivation or organizational identification or by the support provided by the organization to realize long-term goals? In this regard, employee surveys capturing such factors, both before and after implementing a 360-degree system, may prove especially informative. Another interesting question is how much communication and reinforcement of core values is necessary for these values to resonate with—and be remembered by—employees and what are the best ways to do so. We look forward to future research in this area.

Appendix 1: Performance Appraisal Instrument

Core Value #1: **We Gain Control of Our Own Career By Working Hard Every Day and Reaching Out for Support**
We work hard every day to grow and succeed in life, and to make MoPhI successful. We know that by working hard and honestly, we can earn more and we have the chance to be promoted. We reach out to other stores, head office, brands, managers, and the distribution center (DC), to get stock, to get numbers activated, and to get repair and dead on arrival (DOA) cases resolved. By reaching out we achieve higher sales, make more money and have happier customers who will recommend our stores.
Select a number between 1 and 5 for every question where 1 means Never, 2 means Rarely, 3 means Sometimes, 4 means Very often, and 5 means Always.
1.1 Do you explain to all promoters and the cashier their targets and the reward program?
1.2 Do you give all promoters and the cashier their daily target every morning?
1.3 Do you ask all promoters and the cashier every day about their target achievement till date?
1.4 Do you remain positive about targets even if there have been some bad days?
1.5 Do you try to increase sales by reaching out to customers outside the store (for example, by distributing leaflets, making posters of special offers, or telling promoters to stand at the canopy outside the store)?
1.6 Do you work hard to help everyone achieve targets even if popular models are out of stock?
1.7 Do you tell all promoters and the cashier to sell old and stuck models?
1.8 Do you sell to the customer when the brand's promoter is not present?
1.9 Do you make sure any problems in the store (PC, printer, AC, lights, sign board problems, etc.) get fixed?
1.10 Do you stay late if a customer walks into the store at the time of store closing?
1.11 Do you make efforts to get a new promoter if a current promoter resigns or is absent for a long time?
1.12 Do you help get stock from other stores when needed?
1.13 Do you make all the promoters and the cashier believe that they can have a successful career at MoPhI?
1.13b. Do you believe that you can have a successful career at MoPhI?

Appendix 1: Performance Appraisal Instrument (Continuation)

Core Value #2: We Give More Value
We give the best combo offers to our customers. For example, our handset plus headphone offer and our handset plus insurance offer are the best value in the market. We constantly try to learn about the products and services we sell so that we can know what options to offer to our customers and answer their questions better. We help customers with their problems in any way we can. We work to build long term relationships with our customers so that they will visit MoPhI again.
Select a number between 1 and 5 for every question where 1 means Never, 2 means Rarely, 3 means Sometimes, 4 means Very often, and 5 means Always
2.1 Do you know all the DPs, schemes and offers from the different brands (example, Bajaj Finance, EMI scheme, Cash back, PayTM Scheme, etc.)?
2.2 Do you make sure that the price list, posters, and banners in the store are up-to-date?
2.3 Do you know the local market prices?
2.4 Do you teach less experienced team members how to sell profitable bundles to the customer (for example, by giving free gifts, free apps, insurance, unlimited calling, etc. to close sales at a higher price)?
2.5 Do you try hard to close sales against competitors without lowering the price?
2.6 Do you contact the store's customers when the out of stock products become available in the store?
2.7 Do you contact previous customers to tell them about new products?
2.8 Do you make long term relationships with customers?
2.9 Do you ask all promoters and the cashier to make long term relationships with customers?
2.10 Do you make accurate commitments to the customer? (For example, you do not promise that a number will be activated in 3 days, or promise that a handset will be repaired within a certain amount of time if you do not know when it will be repaired) In this question, NEVER means you make promises you do not know MoPhI can fulfill, such as those in the examples, to all of your customers SOMETIMES means you make promises you do not know MoPhI can fulfill, such as those in the examples, to about half of your customers ALWAYS means you never make promises you do not know MoPhI can fulfill, such as those in the examples, to any of your customers
2.11 Do you instruct all promoters and the cashier to make only accurate commitments to the customer?
2.12 Do you help customers that have problems? (Some examples of helping are giving your mobile number to the customer, solving activation problems, showing the customer the way to the service center, or even personally going to the service center with the customer, or sending some person from the store to the service center with the customer) In this question, NEVER means you never take any action to help the store's customers that have problems SOMETIMES means you take one or more actions, such as those described in the examples, to help about half of the store's customers that have problems ALWAYS means you take one or more actions, such as those described in the examples, to help all of the store's customers that have problems
2.13 Do you tell everyone in the store to help customers with problems?
2.14 Do you tell everyone in the store to give the same respect to all customers regardless of their purchase amount? (For example, to be equally respectful to a customer wanting a small Rs. 10 recharge and a customer buying an apple phone)
2.15 Do you tell everyone in the store to be respectful to irritated customers?

Appendix 1: Performance Appraisal Instrument (Continuation)

Core Value #3: We are Honest and Ethical
We are always honest and ethical and we do the right thing at the store every day. We believe that this is the only way to make our store and MoPhI successful.
Select a number between 1 and 5 for every question where 1 means Never, 2 means Rarely, 3 means Sometimes, 4 means Very often, and 5 means Always.
3.1 Are you trustworthy to customers? (Examples of NOT being trustworthy are telling lies about what is being sold to the customer, selling fake products, changing the original batteries of the handset for cheaper batteries, taking the customers' money based on false promises) In this question, NEVER means you take at least one action that is "not trustworthy," such as those described in the examples, with all of the customers you serve SOMETIMES means you take at least one action that is "not trustworthy", such as those described in the examples, with about half of the customers you serve ALWAYS means you never take an action that is "not trustworthy" with any of the customers you serve
3.2 Do you tell all promoters and the cashier to be trustworthy to customers?
3.3 Do you stop wrong activity against the company? (Examples of wrong activities are: stealing, lying, giving unauthorized discounts to friends or family, selling products that are not coming from HO at the store, people making profits for themselves when serving a customer, borrowing store cash or allowing someone to borrow store cash) In this question, NEVER means you never stop wrong activities SOMETIMES means you stop about half of the wrong activities that you notice, such as those described in the examples ALWAYS means you stop all of the wrong activities that you notice, such as those described in the examples
3.4 Do you report wrong activity against the company to HO? Please select "Cannot Answer" if there hasn't been any wrong activity.
3.5 Are you honest at the store? (Being honest means not doing any wrong activity) In this question, NEVER means you do at least one wrong activity, such as those described in the examples, one or more times a day SOMETIMES means you do at least one wrong activity, such as those described in the examples, about once a week ALWAYS means you never do any wrong activity
3.6 Do you tell the cashier and the promoters to be always honest in the store?
3.7 Do you transfer promoters out of your store for personal issues?

Appendix 1: Performance Appraisal Instrument (Continuation)

Core Value #4: We are Caring and Respectful
We care about and respect each other and our customers. We help each other to grow and be more successful. This is who we are.
Select a number between 1 and 5 for every question where 1 means Never, 2 means Rarely, 3 means Sometimes, 4 means Very often, and 5 means Always.
4.1 Do you give the same respect to customers regardless of caste, religion, gender, or economic status?
4.2 Do you give the same respect to the promoters and the cashier regardless of caste, religion or gender?
4.3 Do you earn the trust of the promoters and the cashier?
4.4 Do you tell all promoters and the cashier when they've done a good job (using words such as Well Done, Good job, Keep it up)?
4.5 Do you tell people at the store to work as a team?
4.6 Do you help solve any fights among the promoters or between the cashier and the promoters at the store?
4.7 Do you understand the personal problems of the cashier and the promoters?
4.8 Do you ask for help from HO when a promoter or the cashier needs it?
4.9 Do you care about the promoters' and the cashier's personal development?
4.10 Do you make efforts to learn about new products and services?
4.11 Do you make efforts to learn from the most experienced promoters?
4.12 Do you make efforts to learn from the WhatsApp group?
4.13 Do you try to learn new things?

Appendix 2: Organizational Goals and Values at MoPhI

Vision: To build <u>together</u> the largest and most successful mobile phone retailer in India, providing our people with maximum opportunities for growth

Core values	**Performance metrics used by researchers as proxies of the core values**
We Gain Control of Our Own Career By Working Hard Every Day and Reaching Out for Support	Sales, Gross Profit
We Give More Value	Net Promoter Score, Bundles
We are Honest and Ethical	Abnormal Returns
We are Caring and Respectful	N/A

Appendix 3: Definition of Variables Used in This Study

Variable	Definition
Dependent Variables	
LogSales	Natural logarithm of weekly net sales for store *i* in week *t*
LogGrossProfit	Natural logarithm of gross profit for store *i* in week *t*
BundleInvoice	Percentage of invoices with bundles of products for store *i* in week *t*
NPS	Net promoter score for store *i* in week *t* This variable is constructed based on a question asking customers how likely they are (on a 0 to 10 scale) to recommend MoPhI to a friend. From this question, we calculate the Net Promoter Score as the percentage of "promoters" (respondents giving a 9 or 10) minus the percentage of "detractors" (respondents giving a score of 0 through 6).
AbnormalReturns	Measure of abnormal returns in the first week of the month, calculated as the difference between the average daily returns in the first week of the month and the average daily returns in the rest of the month, scaled by the average daily returns in the rest of the month.
Explanatory Variables	
Post	Indicator variable assuming the value of 1 if week *t* is after the introduction of the 360-degree survey instrument, and zero otherwise
Treatment	Indicator variable assuming the value of 1 if store *i* is in the treatment group, and zero otherwise
StoreManagerChange	Indicator variable assuming the value of 1 if store *i* experienced a change in store manager after the introduction of the 360-degree system.
SalesDays	Number of days the store was open in the week/month (depending on whether the dependent variable is measured at the weekly or monthly level).

Appendix 4: Follow-up interview questionnaire

Question Number	Question
Q1	When did you start working at MoPhI?
Q2.a	Is there anything you like about working at MoPhI? If so, what?
Q2.b	Would you recommend a friend or family member to work at MoPhI?
Q3	Is there anything you don't like about working at MoPhI? If so, what?
Q4	What are your career plans for the future?
Q5	Can you tell me what MoPhI's core values are?
Q6	Do you know what each of these core values mean?
Q7.a	What did you like about the session introducing MoPhI's vision?
Q7.b	Do you have any concerns about the session?
Q8	What do you think is the purpose of the new 360-degree feedback system?
Q9	What is your understanding about who will complete the surveys?
Q10	What is your understanding about who will receive feedback?
Q11	How did you feel about completing a survey about your store manager/yourself for the new 360-degree feedback system?
Q12	What impact, if any, do you think the new 360-degree feedback system will have?
Q13	Has the store manager/Have you discussed the new 360-degree feedback system with your store team?
Q14	(Question for store managers only): How did you feel about others completing a survey about you for the new 360-degree feedback system?
Q15	(Question for store managers only): How do you feel about the upcoming feedback sessions where you will see the results of the surveys and discuss the results with your supervisor?

Appendix 5: John List's 14 tips for pulling off a successful field experiment

1. Use economic theory to guide your design and as a lens to interpret your findings.
2. Be an expert about the market you are studying.
3. Have a proper control group.
4. Obtain sufficient sample sizes.
5. Have a champion within the organization – the higher up the better.
6. Understand organizational dynamics.
7. Organizations that have "skin in the game" are more likely to execute your design and use your results to further organizational objectives.
8. Run the field experiment yesterday rather than tomorrow.
9. Change the nature of the discussion of the cost of the experiment.
10. Make clear that you do not have all the answers.
11. Be open to running experiments that might not provide high-powered research findings in the short run.
12. Don't be captured by the organization.
13. Understand fairness concerns.
14. Always obtain IRB approval.

Appendix 6- Power of the Proposed Tests

We examined the power of our main tests (i.e., those presented in Table 4) to estimate how large the effects of the 360-degree system needed to be, to be detected with 80% power using our data and model specification. We ran power analysis simulations controlling for store fixed effects and clustering standard errors by store, following the methodology described in Bellemare, Bissonnette, and Kroger (2014) and applying this methodology in a similar way as in Li and Sandino (2018). Our simulations assumed the following simplified version of the model used in Table 4:

$$Performance_{it} = \alpha_i + \beta_1 Post_t + \beta_2 Treatment_i * Post_t + \epsilon_{it} \quad (6\text{-}1)$$

where $Performance_{it}$ is the dependent variable (*LogSales, LogGrossProfit, NPS, BundleInvoice, AbnormalReturns*) for store *i* at time *t*, α_i is a representation for store fixed effects, $Post_t$ is an indicator variable for whether time *t* is in the post-treatment period, $Treatment_i$ is an indicator variable for whether store *i* is a treatment store, and β_2 is the treatment effect analyzed.

We used our sample data to estimate how much of the variance of the performance outcomes would be explained by store fixed effects ($var(\alpha_i)$) and how much would remain unexplained ($var(\in_{it})$). As described under "Simulation Details," these inputs were used to obtain each of the points in the power graphs in Figure 6-1, defined based on the following two coordinates:

- β_2 levels (horizontal axis): For each performance outcome variable, we pre-defined different levels of β_2 within a range where the power to detect β_2 increased dramatically. For each β_2 level that we pre-defined, we simulated 100 samples to generate each point in the power graph.
- $Power$ levels (vertical axis): We estimated the power to reject the null $\beta_2 = 0$ at each pre-defined β_2 level following two steps: (a) we ran regressions (using Equation 6-1) for each of the 100 samples simulated for each pre-defined $\beta_2$level, and then (b) we estimated the power to detect that β_2 level based on the fraction of times when the null was rejected at a significance level of 10% and assuming two-sided tests.

Panels A to E of Figure 6-1 show that, with 80% power, we would be able to identify effects of the 360-degree system of at least the following sizes for each performance variable:

Performance Variable	Minimum Detectable Effect Size
LogSales (the natural logarithm of weekly net sales at the store-week level)	16% change in Sales
LogGrossProfit (the natural logarithm of gross profit at the store-week level)	20% change in Gross Profit
NPS (Net Promoter Score: the percentage of "promoters" less the percentage of "detractors")	20% in Net Promoter Score
BundleInvoice (the percentage of invoices including bundles of products)	3% of invoices with bundled products
AbnormalReturns (the abnormal customer daily returns in the first week of every month, scaled by the average daily returns in the rest of the month)	150% in abnormal customer returns

According to our analysis, we would be able to identify a change in net sales due to the 360-degree system of 16% or more relative to the control group and a change in gross profit of 20% or more. We would be able to identify a change in Net Promoter Score due to the 360-degree system of 20% or more (e.g., a shift where 10% of the detractors would become promoters), a change in the incidence of bundled invoices of 3% or more, and a change in abnormal customer returns of 150% or more.

Simulation Details

In our simulation, we generated data as follows:

- $Performance_{it} = \alpha_i + \beta_2 Treatment_i * Post_t + \epsilon_{it}$
- β_2 level was pre-specified
- $Post_t$ was set to one for the second half of the sample period.
- $\alpha_i \sim N(0, var(\alpha_i))$
- $\epsilon_{it} \sim N\big(0, var(\epsilon_{it})\big)$

- 16 of the 32 stores are randomly assigned to the treatment condition, $Treatment_i = 1$ (β_2 is set to the pre-specified level for those stores), the other half is assigned to the control condition, $Treatment_i = 0$.

The following inputs were used to run the simulation:[34]

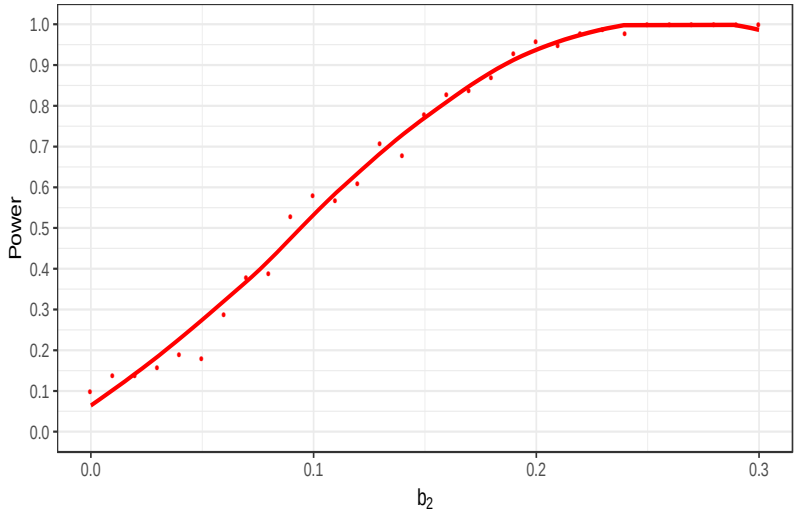
Performance Outcome	# Stores	# Time Periods	# Observations	$var(\alpha_i)$	$var(\epsilon_{it})$
LogSales	32	22	692	1.77	0.16
LogGrossProfit	32	22	688	1.76	0.25
Net Promoter Score	32	22	377	0.04	0.14
BundleInvoice	32	22	692	0.01	0.01
AbnormalReturns	32	6	181	1.21	3.39

For each performance outcome variable and pre-specified level of β_2 used in the data generation process, we generated 100 samples, ran 100 regressions (one per sample), and estimated the power value as the proportion of times when the estimated coefficient $\hat{\beta}_2$ was significantly different from zero. Each point in Figure 6-1 reports the power value for each pre-defined β_2.

Figure 6-1 Power Analysis Graphs

Power to detect an effect size of β_2 on the performance outcomes specified in each panel, accounting to equation (1).

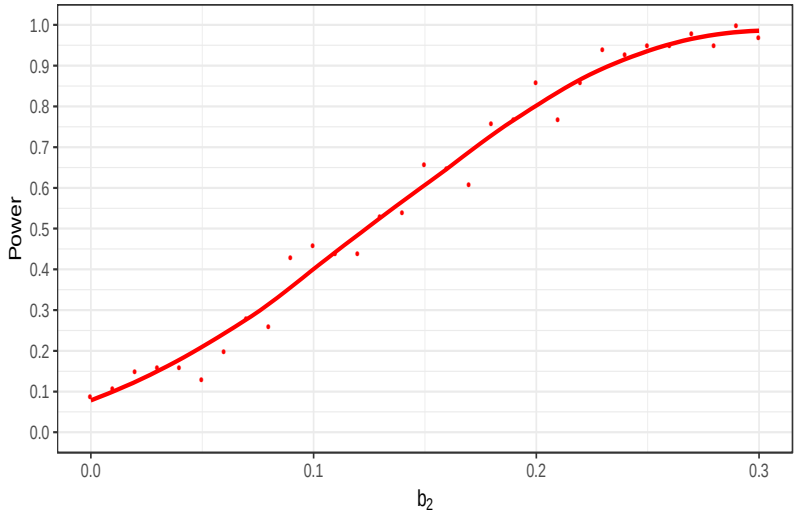
Panel A – Natural Logarithm of Weekly Net Sales



[34] We used a fixed effects regression model including all the explanatory variables that appear in Table 4 to estimate the amount of variance attributed to store fixed effects (var(α_i)) and residual noise (var(ϵ_{it})). Also there are only 6 time periods for the AbnormalReturns variable since it is collected monthly.

Figure 6-1 Power Analysis Graphs (Continuation)

Panel B – Natural Logarithm of Gross Profit



Panel C – Net Promoter Score (the percentage of "promoters" less the percentage of "detractors")

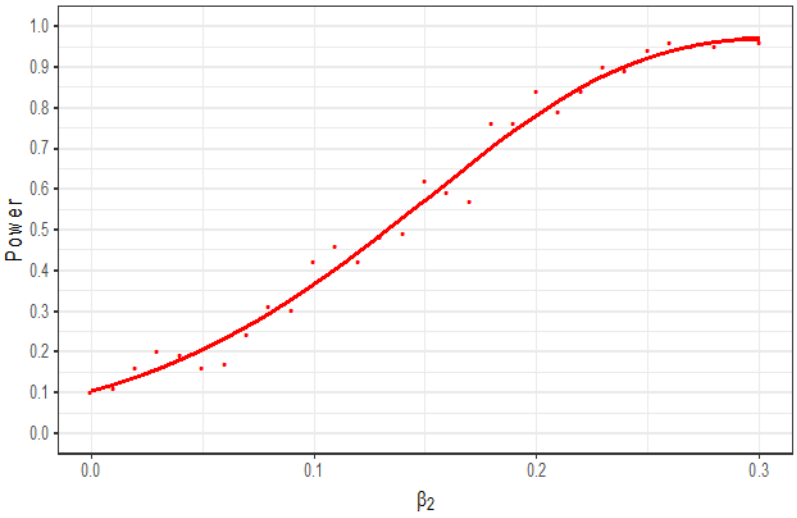
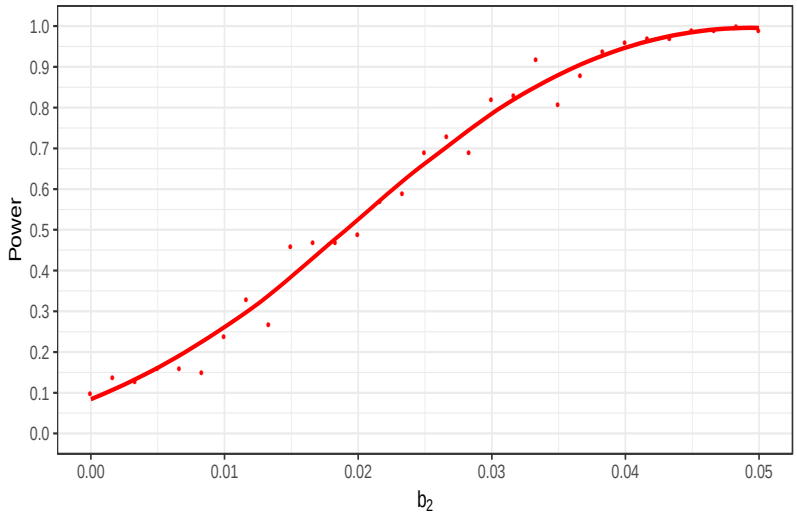
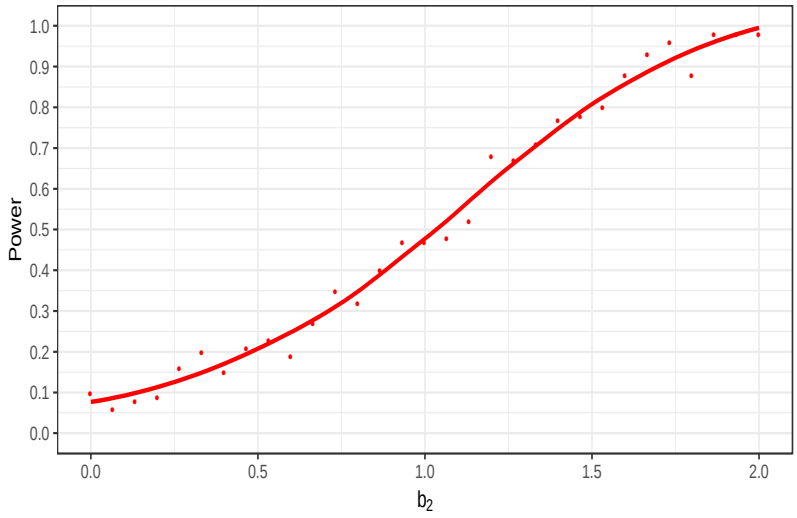


Figure 6-1 Power Analysis Graphs (Continuation)

Panel D – Bundle Invoice (the percentage of invoices including promotion-related bundles)



Panel E – Abnormal Return (the abnormal quantity of customer returns in the first week of every month relative to the normal quantity of customer returns in the other weeks of the month)



References:

Akerlof, G. A., and R. E. Kranton. 2005. Identity and the Economics of Organizations. *Journal of Economic Perspectives* 19(1): 9-32.

Akerlof, G. A., and R. E. Kranton. 2000. Economics and Identity. *The Quarterly Journal of Economics* 115(3): 715-753.

Arnold, T. J., R. W. Palmatier, D. Grewal, and A. Sharma. 2009. Understanding Retail Managers' Role in the Sales of Products and Services. *Journal of Retailing* 85(2): 129-144.

Atwater, L.E., J. F. Brett, and A. C. Charles. 2007. Multisource Feedback: Lessons Learned and Implications for Practice. *Human Resource Management* 46(2): 285-307.

Atwater, L.E. and J.F. Brett. 2005. Antecedents and Consequences of Reactions to Developmental 360° Feedback. *Journal of Vocational Behavior* 66(3): 532-548.

Avery, K. 2000. The Effects of 360 Degree Feedback over Time. *ProQuest Dissertations and Theses.*

Baker, G. P., R. Gibbons, and K. J. Murphy. 1994. Subjective Performance Measures in Optimal Incentive Contracts. *The Quarterly Journal of Economics* 109 (4):1125-1156.

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

Banker, R. D., and S. Datar. 1989. Sensitivity, Precision, and Linear Aggregation of Signals for Performance Evaluation. *Journal of Accounting Research* 27(1): 21-39.

Banker, R. D., G. Potter, and D. Srinivasan. 2000. An Empirical Investigation of an Incentive Plan that Includes Nonfinancial Performance Measures. *The Accounting Review* 75(1): 65-92.

Blanes i Vidal, J., and M. Nossol. 2011.Tournaments Without Prizes: Evidence from Personnel Records. *Management Science* 57(10): 1721-1736.

Bloomfield, R., M. W. Nelson, and E. Soltes. 2016. Gathering Data for Archival, Field, Survey, and Experimental Accounting Research. *Journal of Accounting Research* 54(2): 341-395.

Bol, J. 2011. The Determinants and Performance Effects of Managers' Performance Evaluation Biases. *Accounting Review* 86(5): 1549-1575.

Bracken, D., and W. Rose. 2011. When Does 360-Degree Feedback Create Behavior Change? And How Would We Know It When It Does? *Journal of Business and Psychology* 26(2): 183-192.

Bradler, C. R. Dur, S. Neckermann, and A. Non. 2016. Employee Recognition and Performance: A Field Experiment. *Management Science* 62(11): 3085-3099.

Campbell, D., S. M. Datar, S. L. Kulp, and V. G. Narayanan. 2015. Testing Strategy with Multiple Performance Measures: Evidence from a Balanced Scorecard at Store24. *Journal of Management Accounting Research* 27(2): 39-65.

Campbell, D. 2008. Nonfinancial Performance Measures and Promotion-based Incentives. *Journal of Accounting Research* 46(2): 297-332.

Casas-Arce, P., S. Lourenco, and F. A. Martinez-Jerez. 2017. The Performance Effect of Feedback Frequency and Detail: Evidence from a Field Experiment in Customer Satisfaction. *Journal of Accounting Research* 55(5): 1051-1088.

Charness, G., and U. Gneezy. 2009. Incentives to Exercise. *Econometrica* 77(3): 909-931.

Deller, C. 2018. Reflections on Obtaining Archival Data from the Field. *Journal of Financial*

Reporting, Forthcoming.
Dikolli, S. S., and K. L. Sedatole. 2007. Improvements in the Information Content of Nonfinancial Forward-Looking Performance Measures: a Taxonomy and Empirical Application. *Journal of Management Accounting Research* 19(1): 71-104.
Floyd, E., and J. A. List. 2016. Using Field Experiments in Accounting and Finance. *Journal of Accounting Research* 54(2): 437-475.
Gibbons, R., and K. J. Murphy. 1992. Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence. *Journal of Political Economy* 100(3): 468-505.
Gibbs, M., K. A. Merchant, W. A. Van der Stede, and M. E. Vargus. 2004. Determinants and Effects of Subjectivity in Incentives. *The Accounting Review* 79(2): 409-436.
Griffith, R., and A. Neely. 2009. Performance Pay and Managerial Experience in Multitask Teams: Evidence from Within a Firm. *Journal of Labor Economics* 27(1): 49-82.
Hall, M. 2008. The Effect of Comprehensive Performance Measurement Systems on Role Clarity, Psychological Empowerment and Managerial Performance. *Accounting, Organizations and Society* 33(2-3): 141-163.
Hazucha, J. F., S. A. Hezlett, and R. J. Schneider. 1993. The Impact of 360-degree Feedback on Management Skills Development. *Human Resource Management* 32(2-3): 325.
Holmstrom, B., and P. Milgrom. 1991. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law Economics & Organization* 7(Special Issue): 24-52.
Ittner, C. D., and D. F. Larcker. 1998. Innovations in Performance Measurement: Trends and Research Implications. *Journal of Management Accounting Research* 10: 205-238.
Ittner, C. D., D.F. Larcker, and T. Randall. 2003. Performance Implications of Strategic Performance Measurement in Financial Services Firms. *Accounting, Organizations and Society* 28(7-8): 715-741.
Kaplan, R. S., and D. P. Norton. 1992. The Balanced Scorecard - Measures that Drive Performance. *Harvard Business Review*: 71-79.
Li, S. X., and T. Sandino. 2018. Effects of an Information Sharing System on Employee Creativity, Engagement, and Performance. *Journal of Accounting Research* 56(2): 713-747.
List, J. A., 2011. Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off. *Journal of Economic Perspectives* 25(3): 3-16.
London, M., and R. W. Beatty. 1993. 360-Degree Feedback as a Competitive Advantage. *Human Resource Management* 32(2-3): 353-372.
London, M., and J. Smither. 1995. Can Multi-Source Feedback Change Perceptions of Goal Accomplishment Self-Evaluations, and Performance-Related Outcomes? Theory-Based Applications and Directions for Research. *Personnel Psychology* 48(4): 803-839.
Loughry, M. L., and H. L. Tosi. 2008. Performance Implications of Peer Monitoring. *Organization Science* 19(6): 876-890.
Mael, F., and B. E. Ashforth. 1992. Alumni and their Alma Mater: A Partial Test of the Reformulated Model of Organizational Identification. *Journal of Organizational Behavior* 13(2): 103-123.
Merchant, K. and T. Sandino. 2009. Four Options for Measuring Value Creation. *Journal of Accountancy.* August Issue: 34-37.
Prendergast, C. 1999. The Provision of Incentives in Firms. *Journal of Economic Literature* 37(1): 7-63.

Reilly, R., J. W. Smither, and N. L.Vasilopoulos. 1996. A Longitudinal Study Of Upward Feedback. *Personnel Psychology* 49(3): 599-612.
Sawyer, J. E. 1992. Goal and Process Clarity: Specification of Multiple Constructs of Role Ambiguity and a Structural Equation Model of their Antecedents and Consequences. *Journal of Applied Psychology* 77(2): 130-142.
Smith, A. F. R. and Fortunato, V. J. 2008. Factors Influencing Employee Intentions to Provide Honest Upward Feedback Ratings. *Journal of Business and Psychology* 22(3): 191-207.
Smither, J. W., M. London, N. Vasilopoulos, R. R. Reilly, R. E. Millsap, and N. 1995. An Examination of the Effects of an Upward Feedback Program Over Time. *Personnel Psychology* 48(1): 1-34.
Smither, J. W., M. London, and R. R. Reilly. 2005. Does Performance Improve Following Multisource Feedback? A Theoretical Model, Meta-analysis, and Review of Empirical Findings. *Personnel Psychology* 58(1): 33-66.
Smither, J. W., M. London, and K. R. Richmond. 2005. The Relationship Between Leaders' Personality and Their Recations to and Use of Multisource Feedback. *Group & Organization Management* 30(2): 181-210.
Tayler, W. B. 2010. The Balanced Scorecard as a Strategy-Evaluation Tool: The Effects of Implementation Involvement and a Causal-Chain Focus. *The Accounting Review* 85(3): 1095-1117.
Van den Steen, E. 2010. Culture Clash: The Costs and Benefits of Homogeneity. *Management Science* 56(10): 1718-1738.
van Knippenberg, D. 2000. Work Motivation and Performance: A Social Identity Perspective. *Applied Psychology: An International Review* 49(3): 357-371.
Walker, A. G., and J. W. Smither. 1999. A Five-Year Study of Upward Feedback: What Managers Do with Their Results Matters. *Personnel Psychology* 52(2): 393-423.

Table 1: Descriptive Statistics

Variable	N	mean	sd	p25	p50	p75	min	max
LogSales	692	11.897	1.586	11.046	12.085	13.001	6.040	14.856
LogGrossProfit	688	9.452	1.589	8.698	9.755	10.562	2.565	12.276
NPS	368	0.334	0.394	0.000	0.333	0.553	-1.000	1.000
BundleInvoice	693	0.257	0.129	0.162	0.246	0.345	0.000	0.714
AbnormalReturns	181	0.387	1.925	-0.473	-0.014	0.500	-1.000	16.143
StoreManagerChange	693	0.039	0.194	0.000	0.000	0.000	0.000	1.000
SalesDays (weekly)	692	6.522	1.062	7.000	7.000	7.000	1.000	7.000
SalesDays (monthly)	190	28.200	4.285	28.000	30.000	31.000	12.000	31.000

Notes: In order to have a consistent length in the pre- and post- periods we have restricted the sample to include 11 weeks in the pre-period and 11 weeks in the post-period. Our sample includes store-week level observations for all variables except for *AbnormalReturns*, and *SalesDays (monthly)* which are measured at a monthly level. Observations for the net promoter score (*NPS*) were not available for the entire pre-period, as these metrics were introduced only three weeks before the intervention. All variables are defined in Appendix 4.

Table 2: Correlation Table

		(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
(1)	*LogSales*	1.000							
(2)	*LogGrossProfit*	0.959***	1.000						
(3)	*NPS*	-0.031	-0.029	1.000					
(4)	*BundleInvoice*	0.305***	0.289***	-0.011	1.000				
(5)	*AbnormalReturns*	-0.156***	-0.136**	-0.106**	0.159***	1.000			
(6)	*StoreManagerChange*	0.031	0.029	-0.075	-0.126**	-0.044	1.000		
(7)	*SalesDays (weekly)*	0.551***	0.607***	-0.044	0.240***	-0.168***	0.066	1.000	
(8)	*SalesDays (monthly)*	0.573***	0.627***	-0.073	0.288***	-0.174***	0.070	0.838***	1.000

Notes: Table 2 reports the pairwise Pearson correlation coefficients across all variables of interest in this study. Two-tailed statistical significance is indicated, respectively, with: * = ($p<0.10$); ** = ($p<0.05$); *** = ($p<0.01$). All variables are defined in Appendix 4.

Table 3: Univariate Analyses

		Control		Treatment		Difference (Treatment - Control)	
		Obs	**Mean**	**Obs**	**Mean**	**Obs**	**Mean**
Pre	*LogSales*	176	11.583	176	12.312	352	0.729 ***
	LogGrossProfit	175	9.124	176	9.835	351	0.712 ***
	NPS	40	0.127	43	0.176	83	0.049
	BundleInvoice	176	0.293	176	0.274	352	-0.019
	AbnormalReturns	47	1.058	46	0.445	93	-0.614
Post	*LogSales*	164	11.362	176	12.292	340	0.929 ***
	LogGrossProfit	162	8.932	175	9.874	337	0.943 ***
	NPS	131	0.380	154	0.394	285	0.014
	BundleInvoice	165	0.233	176	0.225	341	-0.008
	AbnormalReturns	42	-0.016	46	0.010	88	0.026
Difference (Post - Pre)	*LogSales*	340	-0.221	352	-0.021	692	0.200
	LogGrossProfit	337	-0.192	351	0.039	688	0.231
	NPS	171	0.252 **	197	0.218 ***	368	-0.034
	BundleInvoice	341	-0.060 ***	352	-0.049 ***	693	0.011
	AbnormalReturns	89	-1.074 **	92	-0.434 *	181	0.640

Notes: Table 3 reports the results of paired t-tests allowing for unequal variances across groups. Rows report comparisons between treatment stores and control stores, while columns report comparisons between pre- and post- periods. We test differences in performance dimensions included in the 360º intervention relative to each of the firm's organizational values. The null hypothesis is that mean(control)=mean(treatment) in the rows, and that mean(pre)=mean(post) in the columns. The results of tests for the alternative hypothesis mean(control)≠mean(treatment) are reported in the rightmost column, and those for the tests of the alternative hypothesis mean(pre)≠mean(post) are reported in the bottom row. The bottom right corner reports the results of paired t-tests relative to the combination of pre/post and treatment/control. Two-tailed statistical significance is indicated, respectively, with: * = ($p<0.10$); ** = ($p<0.05$); *** = ($p<0.01$). All variables are defined in Appendix 4.

Table 4: Effects of the Introduction of the Values-Based 360° Assessment System on Performance

This table reports results from running the following regression:

$$Performance_{i,t} = \alpha + \beta_1 Post_t + \beta_2 Post_t * Treatment_i + \beta_3 StoreManagerChange_{i,t} + \beta_4 SalesDays_{i,t} + \beta_n (Store\ Fixed\ Effects) + \varepsilon$$

Core Values	*We Gain Control of our Career by Working Hard*		*We Give More Value*		*We are Honest and Ethical*
Outcomes	*LogSales*	*LogGrossProfit*	*NPS*	*Bundle Invoice*	*Abnormal Returns*
Post	-0.259**	-0.264**	0.246**	-0.062***	-0.917*
	(-2.50)	(-2.52)	(2.63)	(-3.09)	(-1.87)
*Post*Treatment*	0.211*	0.260*	-0.043	0.013	0.234
	(1.75)	(1.84)	(-0.35)	(0.48)	(0.46)
StoreManagerChange	0.072	0.013	-0.053	-0.021	-0.148
	(0.49)	(0.08)	(-0.63)	(-0.62)	(-0.27)
SalesDays	0.295***	0.279***	0.101*	-0.003	0.230***
	(4.78)	(3.48)	(1.53)	(-0.42)	(2.65)
Intercept	10.040***	7.693***	-0.516	0.302***	-5.842**
	(25.38)	(15.00)	(-1.17)	(7.33)	(-2.42)
Store FE?	YES	YES	YES	YES	YES
N	692	688	368	692	181
adj. R^2	0.208	0.126	0.067	0.121	0.066
adj. R^2 (alt. est.)	0.938	0.899	0.077	0.633	0.123

Notes: We use a difference-in-differences specification and estimate regression coefficients using OLS with standard errors clustered by store. Because the assignment of a store to the treatment versus control group is time-invariant, the inclusion of store fixed effects causes the coefficient relative to the variable *Treatment* not to be estimated, hence we are not reporting a row for this variable in Table 4. In order to maintain the symmetry in the length of the pre- and post-period, having performance data available for only 11 weeks after the introduction of the 360º system. The variable *SalesDays* is expressed as the number of days the store was open for business in the week (month) if the model is estimated to predict a weekly (monthly) dependent variable. In all cases, t-statistics are reported in parentheses underneath the corresponding estimated coefficient. Two-tailed statistical significance is indicated, respectively, with: * = ($p<0.10$); ** = ($p<0.05$); *** = ($p<0.01$). Our estimations are performed using the Stata procedure *xtreg*, with fixed effects and standard errors clustered at the store level. While this procedure yields appropriate standard errors, the R^2 is generally underestimated. In the last row we report the R^2 relative to the estimation of the same model using the Stata procedure *areg*, which fits a linear regression absorbing the categorical factor *Store*, which yields a more realistic R^2 in settings where the number of clusters is large.

Table 5: Supplemental Analyses: Moderation Effect of Store Manager Tenure on the Performance Effects of the Introduction of the Values-Based 360° Assessment System

This table reports results from running the following regression:

$$Performance_{i,t} = \alpha + \beta_1 Post_t + \beta_2 Post_t * Treatment_i + \beta_3 Post_t * Tenure_i + \beta_4 Post_t * Treatment_i * Tenure_i + \beta_5 StoreManagerChange_{i,t} + \beta_6 SalesDays_{i,t} + \beta_n(Store\ Fixed\ Effects) + \varepsilon$$

Core Values	*We Gain Control of our Career by Working Hard*		*We Give More Value*		*We are Honest and Ethical*
Outcomes	*LogSales*	*LogGrossProfit*	*NPS*	*Bundle Invoice*	*Abnormal Returns*
Post	-0.143*	-0.252**	0.424***	-0.058*	-0.019
	(-1.79)	(-2.64)	(3.96)	(-1.93)	(-1.04)
*Post*Treatment*	0.103	0.176	-0.111	0.018	-0.262
	(0.85)	(1.13)	(-0.79)	(0.47)	(0.55)
*Post*Tenure*	-0.001	-0.001	-0.003***	0.000	0.000
	(-0.75)	(-0.69)	(-3.29)	(0.44)	(0.20)
*Post*Treatment*Tenure*	0.001	0.001	0.002**	-0.000	-0.004
	(0.71)	(0.99)	(2.13)	(-0.53)	(-0.80)
StoreManagerChange	0.042	0.046	-0.134*	-0.024	-0.377
	(0.28)	(0.32)	(-1.87)	(-0.64)	(-0.88)
Sales Days	0.184***	0.200**	-0.106	-0.018*	-0.051
	(3.40)	(2.23)	(-0.85)	(-1.84)	(-0.66)
Intercept	11.078***	8.531***	0.845	0.412***	2.008
	(30.31)	(14.06)	(1.02)	(6.10)	(0.91)
Store FE?	YES	YES	YES	YES	YES
N	567	567	313	567	149
adj. R^2	0.079	0.080	0.091	0.113	-0.005
adj. R^2 (alt. est.)	0.940	0.900	0.108	0.684	0.245

Notes: We use a difference-in-differences specification and estimate regression coefficients using OLS with standard errors clustered by store. The variable *Tenure* measures the total number of months of employment recorded for the store manager computed at the time of the intervention. Because the assignment of a store to the treatment versus control group is time-invariant, the inclusion of store fixed effects causes the coefficient relative to the variables *Treatment* and *Tenure* not to be estimated, hence we are not reporting a row for these variable in Table 5. In order to maintain the symmetry in the length of the pre- and post-period, having performance data available for only 11 weeks after the introduction of the 360º system. The variable *SalesDays* is expressed as the number of days the store was open for business in the week (month) if the model is estimated to predict a weekly (monthly) dependent variable. In all cases, t-statistics are reported in parentheses underneath the corresponding estimated coefficient. Two-tailed statistical significance is indicated, respectively, with: * = (p<0.10); ** = (p<0.05); *** = (p<0.01). Our estimations are performed using the Stata procedure *xtreg*, with fixed effects and standard errors clustered at the store level. While this procedure yields appropriate standard errors, the R^2 is generally underestimated. In the last row we report the R^2 relative to the estimation of the same model using the Stata procedure *areg*, which fits a linear regression absorbing the categorical factor *Store*, which yields a more realistic R^2 in settings where the number of clusters is large.

Table 6: Supplemental Analyses: Moderation Effect of Days Sales in Inventory (DSI) on the Performance Effects of the Introduction of the Values-Based 360° Assessment System

This table reports results from running the following regression:

$$Performance_{i,t} = \alpha + \beta_1 Post_t + \beta_2 Post_t * Treatment_i + \beta_3 Post_t * HighestDSI_i + \beta_4 Post_t * Treatment_i * HighestDSI_i + \beta_5 StoreManagerChange_{i,t} + \beta_6 SalesDays_{i,t} + \beta_n(Store\ Fixed\ Effects) + \varepsilon$$

Core Values	*We Gain Control of our Career by Working Hard*		*We Give More Value*	
Outcomes	*LogSales*	*LogGrossProfit*	*NPS*	*Bundle Invoice*
Post	-0.297**	-0.389***	0.251**	-0.061**
	(-2.57)	(-3.05)	(2.37)	(-2.53)
*Post*Treatment*	0.208	0.343**	-0.048	-0.004
	(1.64)	(2.15)	(-0.35)	(-0.13)
*Post*HighestDSI*	0.106	0.352*	-0.015	-0.004
	(0.46)	(1.80)	(-0.07)	(-0.11)
*Post*Treatment*HighestDSI*	0.226	-0.019	0.018	0.131*
	(0.69)	(-0.06)	(0.07)	(1.74)
StoreManagerChange	0.069	0.005	-0.052	-0.021
	(0.49)	(0.03)	(-0.61)	(-0.64)
Sales Days	0.295***	0.278***	0.101	-0.003
	(4.85)	(3.72)	(1.54)	(-0.48)
Intercept	10.043***	7.698***	-0.514	0.304***
	(25.79)	(15.98)	(-1.18)	(7.65)
Store FE?	YES	YES	YES	YES
N	692	688	368	692
adj. R^2	0.215	0.142	0.062	0.152
adj. R^2 (alt. est.)	0.942	0.910	0.067	0.660

Notes: We use a difference-in-differences specification and estimate regression coefficients using OLS with standard errors clustered by store. The variable *HighestDSI* is an indicator assuming value 1 if the individual store fell in the highest quartile of days sales in inventory in the pre-period, and 0 otherwise. Because the assignment of a store to the treatment versus control group is time-invariant, the inclusion of store fixed effects causes the coefficient relative to the variables *Treatment* and *HighestDSI* not to be estimated, hence we are not reporting a row for these variable in Table 6. In order to maintain the symmetry in the length of the pre- and post-period, having performance data available for only 11 weeks after the introduction of the 360º system. The variable *SalesDays* is expressed as the number of days the store was open for business in the week (month) if the model is estimated to predict a weekly (monthly) dependent variable. In all cases, t-statistics are reported in parentheses underneath the corresponding estimated coefficient. Two-tailed statistical significance is indicated, respectively, with: * = ($p<0.10$); ** = ($p<0.05$); *** = ($p<0.01$). Our estimations are performed using the Stata procedure *xtreg*, with fixed effects and standard errors clustered at the store level. While this procedure yields appropriate standard errors, the R^2 is generally underestimated. In the last row we report the R^2 relative to the estimation of the same model using the Stata procedure *areg*, which fits a linear regression absorbing the categorical factor *Store*, which yields a more realistic R^2 in settings where the number of clusters is large.