DO MANAGEMENT INTERVENTIONS LAST?
EVIDENCE FROM INDIA

Nicholas Bloom[a], Aprajit Mahajan[b], David McKenzie[c] and John Roberts[d]

June 20, 2018

Abstract:

Beginning in 2008, we ran a randomized controlled trial that changed management practices in a set of Indian weaving firms (Bloom *et al*. 2013). In 2017 we revisited the plants and found three main results. First, while about half of the management practices adopted in the original experimental plants had been dropped, there was still a large and significant gap in practices between the treatment and control plants. Likewise, there remained a significant performance gap between treatment and control plants, suggesting lasting impacts of effective management interventions. Second, while few management practices had demonstrably spread *across* the firms in the study, many had spread *within* firms, from the experimental plants to the non-experimental plants, suggesting limited spillovers between firms but large spillovers within firms. Third, managerial turnover and the lack of director time were two of the most cited reasons for the drop in management practices in experimental plants, highlighting the importance of key employees.

JEL No. L2, M2, O14, O32, O33.

Keywords: management, organization, productivity, and India.

Acknowledgements: Financial support was provided by SEED at the Graduate School of Business at Stanford, by the Stanford Center for Poverty and Development, and by the World Bank under the Strategic Research Program (SRP). This research would not have been possible without the consulting team of Saurabh Bhatnagar, Shaleen Chavda, Rahul Dsouza, Sumit Kumar, and Ashutosh Tyagi. We thanks our formal discussant Rebecca Henderson, Larry Katz and seminar participants at Duke, IGC, Maryland, NBER, Stanford and the World Bank for comments. AEA RCT Registry identifying number AEARCTR-0002808.

[a] Stanford Economics (nbloom@stanford.edu); [b] UC Berkeley Agricultural and Resource Economics (aprajit@gmail.com); [c] Development Research Group, The World Bank (dmckenzie@worldbank.org); [e] Stanford Graduate School of Business (jroberts@stanford.edu)

I. INTRODUCTION

After an early recognition of management as a driver of differences in firm performance (e.g. Walker, 1887 and Marshall, 1887), economists are again paying increasing attention to the role of management in firm and economy-wide performance (Roberts, 2018). Whereas the size and profitability of the management consulting industry is often cited as a revealed preference measure of the importance of management, recent academic work has also established a credible causal link between changes in management practices and productivity in medium and large firms (Bloom *et al*, 2013; Bruhn *et al*, 2018). The longer-term persistence of management improvements caused by consulting interventions, however, remains an open question. [1] The received wisdom at a leading global management consulting firm when two of the authors were employed there was that such innovations lasted approximately three years.

Competing views of management offer differing predictions about the persistence of consulting-induced improvements in management practices. One view, best exemplified by the "Toyota way" (Liker, 2004) views management improvements as launching a continuous cycle of improvement, as systems put in place for measuring, monitoring, and improving operations and quality enable constant improvement. A related idea is that management practices are complementary to one another, so that the costs of adding new practices fall as others are put in place. For example, in our context of cotton weaving, scientific management of inventory levels will only be possible once the firm has put in place systems to record all yarn transactions and to regularly monitor stock levels.

A countervailing view argues that maintaining good management is difficult, with many of the companies extolled in business books as paragons of good management subsequently failing (*The Economist*, 2009, Kiechel, 2012). This may be even harder when changes are introduced externally, with the Boston Consulting Group (BCG) reporting that two-thirds of transformation initiatives ultimately fail (Sirkin *et al*, 2005). This finding presumably refers to high-level strategic and organizational change efforts in large firms that would use BCG. But both Karlan *et al*. (2015)

[1] To our knowledge Giorcelli (2017), who uses observational data to examine the effect of Marshall Plan sponsored management training on long-term outcomes, is the only other work that examines persistence in a causal framework.

and Higuchi *et al*. (2016) find that light consulting engagements in smaller firms than the ones we studied led to firms' gradually discarding practices over the subsequent three years. One reason may be that these practices are inappropriate and will be abandoned as firms learn that they are not suitable in their setting.

This paper examines the persistence of management practices adopted after an extensive, consultant-supported intervention that we undertook in a set of multi-plant Indian textile weaving firms from 2008 to 2010 (see Bloom *et al*, 2013 for a more detailed description). The intervention took the form of a randomized controlled trial. Firms were randomly allocated into treatment and control groups, and the intervention was done at the plant level within each firm. Both treatment and control plants were given recommendations for improving management practices in several areas, and the treatment plants received additional consulting help in implementing the recommendations. The intervention led to a substantial uptake of the recommended practices in the treatment plants and a modest one in the control plants, with corresponding improvements in various measures of performance.

We stopped observing the firms in 2011, but we wondered – as did many in our audiences – about whether these changes would last. As a result, we returned to the study firms in 2017 with the same consulting team and collected data on management practices and basic firm performance. We found that both treatment and control experimental plants had in fact dropped some practices, though fewer than we and the consultants had forecast. Since the control plants also dropped practices, the treatment effect on practices is constant over time, at 20 percentage points. Meanwhile, the plants in the treatment firms that had not been part of the experiment (treatment firms typically had multiple plants) had adopted many of the recommendations, so their package of current practices were very close to those of the treatment plants.

We were also able to collect information on the reasons for the dropping of management practices. We find that practices are more likely to be dropped when the plant manager changes, when the directors (the CEO and CFO) are busier, and when the practice is one that is not commonly used

in many other firms. The first two reasons highlight the importance of key employees within the firm for driving management practices,[2] while the latter emphasizes the importance of beliefs.

Although budgetary constraints rendered us unable to measure long-term impacts on firm profits or overall productivity, we are able to track changes in looms per worker, a simple and commonly-used proxy for labor productivity in the industry, and use this to impute worker productivity. Despite their dropping some practices, we find treated firms show lasting improvements in worker productivity, which is 35% higher than in the control group after 8 years; that treated firms have gone on to use more consulting services of their own accord; and that they have supplemented the operational management practices introduced by the consultants from our study with better marketing practices.

This paper is related to several literatures, including the drivers of firm and national productivity (see, e.g., Syverson 2011), on management randomized control trials (see, for example, Anderson *et al*. 2017; McKenzie and Woodruff 2014) and the large literature on the importance of management for firm performance (e.g. Osterman 1994, Huselid 1995, Ichniowski *et al*. 1997, Capelli and Neumark 2001, Braguinsky *et al*. 2015, and Fryer 2017). Section II of the paper discusses the original consulting experiment, section III describes the follow-up and section IV offers concluding remarks.

II. THE 2008-2010 CONSULTING EXPERIMENT

II.A. The Experimental Design

Our original experiment measured the impact of improving management practices in a set of large textile firms near Mumbai in 2008. The experiment involved 28 plants across 17 firms in the woven cotton fabric industry. These firms had been in operation for 20 years on average, and were family-owned and managed. They produced fabric for the domestic market (although a few also exported). Table 1 reports summary statistics for the textile manufacturing parts of these firms (a few of the firms had other businesses in textile processing, retail and real estate). On average the study firms

2 This links to the literature on management and CEOs – for example, Bertrand and Schoar (2003), Bennesden et al. (2007), Lazear et al. (2016) and Bandiera et al. (2017).

had about 270 employees, assets of $8.5 million and annual sales of $7.5 million. Compared to US manufacturing firms, these firms would be in the top 1% by employment and the top 4% by sales, and compared to Indian manufacturing firms they are in the top 1% by both measures (Hsieh and Klenow, 2010). Hence, these are large manufacturing firms.[3]

These firms are complex organizations, with a median of 2 plants per firm (in addition to a head office in Mumbai) and 4 reporting levels from the shop-floor to the managing director. The managing director was the largest shareholder in all firms, and all directors belonged to the same family. Two firms were publicly listed on the Mumbai Stock Exchange, although more than 50% of the equity in each of these was held by the managing family.

The field experiment aimed to improve management practices in the treatment plants and we measured the impact of doing so on firm performance. We contracted with a leading international management consultancy firm to work with the plants as the easiest way to change plant-level management practices rapidly. The full-time team of (up to) 6 consultants had been educated at leading Indian business and engineering schools and most of them had prior experience working with U.S. and European multinationals.

The intervention ran from August 2008 until August 2010, with data collection continuing until November 2011. The intervention focused on a set of 38 management practices that are standard in American, European, and Japanese manufacturing firms and which can be grouped into five broad areas: factory operations, quality control, inventory control, human-resources management, and sales and orders management (for details see Appendix Table A1). Each practice was measured as a binary indicator of the adoption (1) or non-adoption (0) of the practice. A general pattern at baseline was that plants recorded a variety of information (often on paper sheets), but had no systems in place to monitor these records or use them in decisions. For example, 93 percent of the treatment plants recorded quality defects before the intervention, but only 29 percent monitored them daily or by the particular sort of defect, and none of them had any standardized system to analyze and act upon this data.

The consulting intervention had three phases. The first phase, called the *diagnostic* phase, took one month and was given to all treatment and control experimental plants. It involved evaluating the current management practices of each plant and constructing a performance

3 Note that most international agencies define large firms as those with more than 250 employees.

database. At the end of the diagnostic phase the consulting firm provided each plant with a detailed analysis of its current management practices and performance and, crucially, recommendations for change.

The second phase was a four-month *implementation* phase given only to the treatment experimental plants. In this phase, the consulting firm followed up on the diagnostic report to help introduce as many of the 38 management practices as the plants could be persuaded to adopt. The consultant assigned to each plant worked with the plant management to put the procedures into place, fine-tune them, and stabilize them so that employees could readily carry them out.

The third phase was a *measurement* phase, which lasted until November 2011. This involved collection of performance and management data from all treatment and control plants. In return for this continuing data, the consultants provided light consulting advice to the treatment and control plants (primarily to keep them involved).

II.B. The Initial Experimental Results – Management Practices

The intervention led to increases in the adoption of the 38 management practices in the treatment plants by an average of 37.8 percentage points by August 2010 (approximately one year after the start of the intervention). This adoption rate dropped by only 3 percentage points in the subsequent year, showing considerable persistence in practices after the consultants had exited the firms. Not all practices were adopted equally, with firms adopting the practices that (unsurprisingly) were the easiest to implement and/or had the largest perceived short-run pay-offs, e.g. the daily quality, inventory and efficiency review meetings. This adoption also occurred gradually, in large part reflecting the time taken for the consulting firm to gain the confidence of the firms' directors. Initially many directors were skeptical about the suggested management changes, and the intervention often started by piloting the easiest changes around quality and inventory in one part of the factory. Once these started to generate improvements, these changes were rolled out and the firms then began introducing the more complex improvements around operations and human resources.

In contrast, the control plants, which were given only the one-month diagnostic and corresponding recommendations, increased their adoption of the management practices, but by only 12 percentage points on average. This is substantially less than the increase in adoption among the treated plants, indicating that the four months of the implementation phase were important in

changing management practices. Table 2 Column 2 reflects this and shows a statistically significant 25 percentage point treatment effect on management practices in 2011. We note that the change for the control firms is still an increase relative to the rest of the industry cluster around Mumbai (which had more than 100 non-project plants), which did not change their management practices on average between 2008 and 2011.

Finally, since these are multi-plant firms and the consulting firm worked at the plant level, the treatment and control firms also had plants that were not part of the intervention, which we label "non-experimental plants." For example, if a treatment firm has three plants A, B and C and the diagnostic and implementation intervention was performed on plant A this would be a "Treatment Experimental plant" while plants B and C would be "Treatment Non-Experimental plants". Likewise if a control firm had plants D, E and F and the diagnostic intervention was only performed on plant D, then D would be an "Control Experimental plant" while E and F would be "Control Non-Experimental plants". Appendix Table A2 reports the breakdown of the plant count into these four groups.

Although the consulting firm did not provide consulting services to the non-experimental plants, it was still able to collect bi-monthly management data and some basic data for these plants. The non-experimental plants in the treatment firms saw a substantial increase in the adoption of management practices. In these 5 plants the adoption rates increased by 17.5 percentage points by August 2010, without any drop in the second year. This increase occurred because the executives of the treatment firms copied the new practices from their experimental plants over to their other (non-experimental) plants. Interestingly, this increase in adoption rates is similar to the control firms' 12 percentage point increase, suggesting that the copying of best practices across plants within firms can be as least as effective at improving management practices as short (1-month) bursts of external consulting.

II.C. The Initial Experimental Results – Firm Performance

Experimental treatment plants experienced a significant increase in output of 9.4% relative to the experimental control plants, which came about both by decreasing quality defects (so that less output was scrapped); and by undertaking routine maintenance of the looms, collecting and monitoring breakdown data, and keeping the factory clean, which reduced machine downtime. Total factor productivity (TFP) increased by 16.6% due to both the increase in output and a

reduction in inputs due to reduced inventory and reduced labor inputs for mending defective fabric. These improvements were estimated to have increased profits per plant by about $325,000 per year. We estimate that this represented, on average, a 130% one-year return on the market cost of the consulting services.

III. THE 2017 FOLLOW UP

III.A. The Follow-up Process

In January 2017, working with the same consulting firm, we re-contacted the 17 textile firms from the original study. Fortunately, all 17 firms agreed to work with the research team again on a follow-up study. This 100% uptake was aided by a combination of three factors: (A) the positive impact of the intervention in the first wave on the firms' management and performance; (B) the stability of the firms, which had maintained the same address and contact details, and (C) the engagement of the same three consulting company partners and project manager as the 2008-2011 intervention.[4] One complication is that one single-plant. treatment firm was in the midst of closing down after the owner's death. Without any close male relatives to continue the business, the owner's widow had decided to sell the business, which, given its location, meant the business would stop trading and the site would be converted into residential housing.[5]

One weakness of this follow-up wave is that our budget allowed us only two months of the consultants' time, which was sufficient to collect management data for all production sites and a basic set of firm performance indicators (e.g. on employment and looms), but not to collect detailed weekly output data that would allow TFP estimation, because that would have required extracting data on a firm-by-firm basis from log-books and accounting software. Consequently, our analysis is confined to management practices and basic performance indicators like employment or looms per employee, along with an imputed measure of labor productivity.[6]

---

4 These personal contacts are very important in our context. In fact, we delayed the start of this project to ensure we could staff the project with the same senior consulting team as the 2008-2011 wave.

5 The firm was over 30 years old, and due to the expansion of Mumbai was now located in a residential area so the land was more valuable as housing than for production.

[6] We were also unable to pursue other, more qualitative, approaches because of the lack of resources.

This follow-up data collection corresponds to an average period of 9 years since the implementation phase of the consulting intervention started and 7 years since it ended. It therefore enables us to examine the long-term persistence of these large changes in management practices.

III.B. Results on Management Practices

In Figure 1 we plot the management scores over time after re-visiting the plants in January 2017 evaluated on the same 38-management practice scoring grid as in the prior experiment. We find substantial persistence of the management intervention, which we summarize below with four main results.

Treatment Experimental Plants: First, the management scores in the treatment plants fell from 0.60 at the end of the last wave to 0.46 eight years later. This drop of 0.14 points in the management score reverses 40% of the original 0.35 increase (noting these firms started pre-treatment with an average management score of 0.25) over an eight-year period. This fall in the management practice score is equivalent to about an annual depreciation rate of 6% in the original increase in management practices.

Control Experimental Plants: Second, the control plants also saw a drop in their management scores, falling by 0.08 points from 0.40 at the end of the last wave to 0.32. This is smaller in absolute terms compared to the fall in scores in the treatment plants, but the increase in management practices in the control plants was only 0.12 points (from an original score of 0.28), so that the drop in practice scores is 66% of the intervention gain, implying about a 13% depreciation rate of the original management increase.

Together this indicates that, even eight years after the initial intervention the treatment firms still had higher management practices. Table 2 reports the results from running the Ancova specification for plants (i) at time (t):

$$\text{Management}_{i,t} = a + b_1*\text{Treatment}_i*\text{Year=2011} + b_2*\text{Treatment}_i*\text{Year=2017} + c*\text{Management}_{i,2008} + e_{i,t}$$

Indeed, we see that the long-run treatment effect in 2017 of 19.7 percentage points is similar in magnitude to the short-run effect in 2011 (20.6 percentage points), and we cannot reject equality of these treatment effects over time (p=0.802). These effects are individually statistically significant both using conventional (large-sample normality-based) inference as well as permutation procedures with exact finite sample size (the corresponding p-values are also reported in Table 2). Thus, the intervention generated persistent impacts on the treatment plants. Moreover, the greater percentage depreciation of the improvements in the control plants (66%) versus the

treatment plants (40%) suggests that small improvements in management may be less stable than large improvements. One possible reason which we discuss further below is that bundles of management practices are complementary, so that adopting only parts of them may be less stable than adopting all of them. Of course, given the small sample sizes in this experiment this could also reflect sampling noise - something that should be remembered when evaluating all our results from this experiment.

Non-experimental plants: Third, the non-experimental plants in the treatment firms experienced a slight improvement in their management practice adoption rates, from 0.43 in 2011 to 0.47 in 2017. Indeed, by 2017 their management scores were very similar overall to the treatment experimental plants (indeed slightly higher, although not significantly so). Similarly, in the control firms the non-experimental plants also converged with the experimental plants (again slightly higher but not significantly). This suggests (as we discuss further below) that the practice improvements in the experimental plants spilled over to the non-experimental plants during the seven years after the intervention.

Expectations on durability of the intervention: Finally, before we re-contacted the firms in December 2016, each member of the consulting team from the original intervention and the academic team provided predictions for the management scores we expected to find on revisiting the firms in 2017.[7]

These expectations were informed by the contrasting views of management improvements noted in the introduction: under the "Toyota way" of continuous improvement we would expect the management practices to not only persist, but to continue to improve in treatment plants so that the gap with the control plants would widen; whereas under the "inappropriate technology" view, we would expect many practices to be dropped and the treatment group to converge back to the control group. The average values of the estimates of the seven team members are shown for the treatment experimental, treatment non-experimental, control experimental plants and control non-experimental plants with the symbols TE, TN, CE and CN respectively on the graph.[8] These

7 Other examples of getting experts to provide ex ante predictions of the results of an experiment can be found in Hirschleifer et al. (2016), Groh et al. (2016) and Dellavigna and Pope (2017).

8 The predictions of the individual consultant and academic team members were made independently – Bloom estimated first and then the other team members individually e-mailed him

predicted values are all below the actual outcomes, indicating that the project team expected steeper declines in management practices relative to what actually occurred, particularly for the non-experimental plants. While some of the practices were dropped, the majority of the interventions remained in place eight years later and the gap with the control group remained steady. The results therefore lie between these two extreme views of constant improvement and of no long-run impact.

To delve further into the management changes, we also analyzed the 38 individual practices as highlighted in Figure 2, which plots the average score for the experimental plants in the treatment firms on each practice on the X-axis against the average scores for the non-experimental plants (in the same firms) on the Y-axis, for the years 2008 (pre-intervention), 2011 (post-intervention) and 2017 (long-run follow-up). We observe that initially the experimental and non-experimental plants in the treatment firms had similar practice scores, with a correlation of 0.91. After the intervention, the scores for the experimental plants improved considerably, leading to an eastward shift in the points and a drop in the correlation to 0.81 (top-right figure). Finally, in the bottom left figure we see the experimental plants and non-experimental plants again have very similar scores (correlation of 0.91), with a reversion of the scores towards the 45-degree line.

Figure 3 complements this by showing the long-difference of management practices in the experimental and non-experimental plants (in the treatment firms) between 2008 and 2017 (left-panel) and 2011-2017 (right panel). This shows that, first, that between 2008 and 2017 both sets of plants adopted similar bundles of management practices. But, second, looking at 2011-2017 we see the timing of these practice adoptions were not the same. The experimental plants adopted most of these practices between 2008-2011, so that from 2011 to 2017 they mostly had negative practice changes. The non-experimental plants, in contrast, were still heavily adopting a number of practices post 2011, so they show a balanced mix of drops and additions post 2011.

So, in summary, Figures 1 to 3 paint a picture of the treatment (and to a lesser extent the control) experimental plants adopting a slew of management practices during the initial intervention phase in 2008-2010, so by 2011 they have substantially higher management scores. These scores subsequently subside as some practices are dropped. The non-experimental plants adopted fewer practices in 2008-2010 but continued to adopt practices, and by 2017 had

---

their predicted scores. The average predicted scores were not particularly different across the two groups (hence we present them averaged together).

comparable scores with the experimental plants. Thus, by 2017 the management practice improvements appear to have equalized over across plants within treatment firms.

III.C. What Drives Changes in Management Practices

We next explore the proximate causes for the adoption or non-adoption of management practices on a practice-by-practice basis in Table 3 using directors' and plant managers' stated reasons for adding or dropping practices. In the "Treatment experimental" column we report the percentage of practices added (top panel) and dropped (bottom panel). In the second, third and fourth panels we report similar figures for the "Treatment non-experimental", "Control Experimental" and "Control Non-experimental", while reporting all plants in the final column. A few results are worth noting.

First, we see that, while a significant fraction of practices remains unchanged from 2011, there is notable churn in management practices across all plants. In particular, 4.1% of practices have been added and 12.4% of practices dropped since the end of the experiment. We are reasonably confident that these are accurately measured, derived as they are from detailed interviews with firm directors and plant managers combined with lengthy firm visits by the consulting team. Second, in the non-experimental plants in the treatment firms, spillovers from other plants (in the same firm) is the single largest reason for practice adoption and accounts for 4.2% of improvements (out of a total improvement rate of 6.9%). In the control firms, spillovers from other firms outside the experimental group[9] were the most important driver of management improvements, driving 2.2% on average of the practice upgrades (out of a total of 2.6%). These two figures highlight the importance of within and across firm spillovers in improving management practices over the long run.

Third, in the experimental plants (in the treatment firms) the major reason for dropping practices was the introduction of a new plant manager (9.9% out of a total of 16.7%, so well over a half). The plant manager was evidently a critical part of the management improvement in the

9 Qualitatively these improvements appear to be from copying other firms in the industry, outside of those in our experimental sample. We did not come across cases of the control firms saying they had learned from the treated firms.

intervention plants, and if he left the firm then many of the practice improvements subsequently collapsed.[10] Another major factor across all the plants was director time – overall 3.6% of practices were dropped when directors had to reduce the time they spent managing the plant, often because of other business commitments (e.g. finance, marketing, or other businesses like retail or real-estate). This highlights the importance of CEO time for firm management, consistent with the work of Bandiera *et al*. (2017). Finally, we see that 4.2% of practices were dropped because of "perceived negative benefits," which means the firms decided the practices were actually not worth adopting and decided to drop them.

Table 4 analyzes the drivers of the changes in management practices by looking at each practice-by-plant cell between 2011 and 2017 in a regression format. Hence, we examine the change in each practice (-1, 0 or 1) for each plant between 2011 and 2017 (for plants present in both years). In column (1) we see the constant term of -0.083 indicates that, on average across plants (experimental and non-experimental plants in treatment and control firms) and practices, the average practice dropped by 8.3% over this period. In column (2) we control for experimental plant status and see this accounts for all the drop, highlighting that management practices scores were roughly constant after 2011 in the treatment non-experimental plants. In column (3) we instead add a treatment dummy and find this is completely insignificant – as can be seen from Figure 1 on average treatment firms did not change (treatment experimental plants dropped their management score and treatment non-experimental plants increased their management score). In column (4) we control for having a new-manager,[11] split this by treatment and control, and see for treatment plants a large significant negative effect (which is driven by the treatment experimental plants) with nothing significant for control plants. This highlights the role of managerial turnover in the drop in management practices in well managed plants. Moreover, presumably, given that

10 See also Fryer (2017) who argues that principal turnover was the primary reason for declines in school performance improvements following an experimental intervention aimed at changing school management practices in the United States.

11 Note that the results in all columns that use post-treatment measured regressors are non-experimental in nature. We do, however, test whether having a new plant-manager is differential across treatment and control, experimental or non-experimental, or correlated with management score in 2011, and find no significant difference. The point-estimate (standard-errors clustered at the firm-level) are 0.050 (0.234), 0.086 (0.222), 0.654 (0.517) respectively. Of course, we should as always be cautious of inference given the small sample size. We were also unable to track down departed plant managers.

management practices will have only recently improved in the experimental plants, they are particularly susceptible to managerial turnover as good practices may not have had time to become established norms.

In column (5) we focus instead on the correlation of changes in practices with the frequency of usage across all plants of the practices in 2008, which is valued from 0 to 1, measuring the share of plants in the pre-experimental period that had adopted this practice. This proxies for how widespread their adoption was prior to the intervention, and the positive coefficient indicates that common practices were more likely to be maintained (so uncommon practices were more likely to be dropped). This highlights that the intervention was more successful at getting badly managed plants to adopt relatively standard practices – such as basic measurement systems – than getting plants to adopt more advanced practices like data review meetings and performance rewards. In column (6) we add these all together and the results look similar, suggesting these are reasonably independent relationships.

Finally, in column (7) we include the management score in 2011 to look for mean reversion, finding a negative but insignificant coefficient. This is confirmed in Figure 4 which shows that both the initial treatment increase in management practices from 2008 to 2011 and the subsequent drop are uncorrelated with initial levels of management practices. So, changes in management practices appear not to be strongly correlated with initial levels, implying that, like TFP, a highly persistent auto-regressive (or random-walk) form of stochastic evolution. Figure 4 is also useful in showing the distribution of changes in management practices among treated plants. We see that every single treated experimental plant improved its practices between 2008 and 2011, and every one of these plants subsequently saw a drop in its management practice score between 2011 and 2017. It is therefore not the case that there were some treated experimental plants in which a "Toyota way" virtuous cycle of continuous improvement occurred.

Finally, we examine the practices that were adopted to see which were the least likely to be retained, and which were the stickiest. Table A3 reports the number of firms which ever adopted a practice (i.e. were not using it in 2008, and then used it in at least one of 2011 or 2017), the number who after adopting were no longer using the practice in 2017, and the proportion of adopters who dropped the practice. We see two types of practices that were most likely to be dropped. The first are a set of visual displays and written practices that very few firms were using before the intervention and then were discarded afterwards. These include displaying written

procedures for warping, drawing, weaving and beam gaiting; displaying standard operating procedures for quality supervisors; and displaying visual reports of daily efficiency by loom and weaver. The second set of practices most likely to be dropped were ones that required daily attention from management: monitoring defects on a daily basis; meeting daily to discuss quality defects and gradation; and updating visual aids of efficiency on a daily basis. They were thus costly, and presumably seen as not very valuable.

In contrast, we see that many of these practices are very sticky. Of our 38 practices, once adopted, 14 are not dropped by a single plant, and a further 8 are dropped by at most one-quarter of adopters. Particularly noticeable among these sticky practices are that those which were adopted by 10 or more plants and then never dropped. These relate very closely to the most immediate improvements in quality and inventory levels that we saw from the original consulting intervention: recording quality defects in a systematic manner (defect-wise); having a system for monitoring and disposing old stock; and carrying out preventative maintenance. Finally, we note that not all daily activities were susceptible to being dropped, with those most closely tied to keeping machines running quite persistent: firms still maintained daily monitoring of machine downtime and had daily meetings with the production team.[12]

III.D. Results on Long Run Performance

The other question we investigated when returning to the plants was the long-run performance impact of the original management interventions. Because of budget limitations and the reluctance of firms to share financial data, we are not able to undertake a detailed analysis of TFP.[13] We were able, however, to collect basic information on plant size and looms in 2014 and 2017 to supplement our original data for 2008 and 2011. Since there were changes over time in the number of plants per firm, and the management practices have converged across plants within firms, we examine performance at the firm level.

---

[12] Breaking down the adoption status by the treatment and experimental status (e.g. "Treatment Non-Experimental Plant") reveals that Control Non-Experimental plants were the least likely to adopt any practices but conditional on adoption did not disadopt them subsequently.

[13] In our original study the consulting firm spent many months extracting production data from firms' log books and production records, which were used to construct a measure of TFP. We were not able to extract this data in our longer-term follow-up.

We run Intention to Treat (ITT) panel regressions over four years (2008, 2011, 2014 and 2017) at the firm level with firm and year fixed effects and standard errors clustered at the firm-level:

$$OUTCOME_{i,t} = aTREAT_{i,t} + b_t + c_i +_{i,t}$$

where OUTCOME is one of the key outcome metrics of looms, looms/employee, etc. We report statistical significance using both conventional inferential procedures based on normal approximations as well as using permutation tests that have exact finite sample size to allay sample size concerns.[14]

We start in column (1) of Table 5 in the top panel looking at the number of looms (in logs), which is a basic measure of production capacity. In panel A, we regress this on a dummy for the year being greater or equal to 2011 - a post-intervention dummy - finding a statistically insignificant coefficient of -0.032. In panel B, we break down this impact by year, with the point estimates suggesting a 16.1 percent increase in capacity by 2017, but this is also not statistically significant.

In column (2) we examine employment. The point estimates suggest a relatively large drop in employment, of 23 to 24 percent on average over the full period, and in 2017. However, this drop is also not statistically significant. There are two reasons why employment may have fallen. The first is that, at baseline, firms employed many workers fixing quality defects and would need less of this sort of labor as quality improved. Second, production process improvements and fewer breakdowns can enable the same worker to be in charge of more looms.

Column (3) combines these measures to focus on our main measure of long-term firm productivity, which is log looms per employee. This is a classic productivity measure in the literature (see, for example, Clark 1987 or Braguinsky *et al*. 2015). One reason is that employees spend much of their time dealing with malfunctioning looms, so that a higher number of looms per employee indicates fewer breakdowns and higher rates of production uptime (the time the loom is producing output rather than being repaired). As such, column 3, panel A, shows that the average treatment effect over the full post-intervention period was to increase looms per employee by a statistically significant 26.7%. Panel B suggests this efficiency improvement was rising over time in that the coefficients are generally larger for 2017 that 2011, with the long-run impact a statistically significant 51.0 percent increase in this productivity measure (the standard errors

[14] We also estimate the regression at the plant level and the results are qualitatively similar.

remain relatively unchanged over time). However, despite the trend of rising point-estimates we cannot reject the null that the productivity impact is constant over time.[15]

We next investigate the impact on labor productivity. While we did not collect information on labor productivity in 2017, we can use the survey data from the initial wave to impute a labor productivity impact. In particular, we use data from a survey we ran in 2011 of 113 firms in the broader textile industry cluster around Mumbai (see details in Appendix A2), in which we collected data on physical production, employment, and looms. Using this, we show in Appendix Table A4 and Figure A2 that there is a strong correlation between labor productivity (output per worker), and looms per worker in both the cross-section and the panel. Taking the fitted coefficient of 0.734 from column (4) of Table A4, we impute labor productivity from looms per employee for our experimental firms. The average imputed increase in labor productivity since 2011 is then 19.0% (exp(0.237*0.734)), and the long-run impact is 35.3% (exp(0.412*0.734)). These impact figures are remarkably similar to the 15.3% and 31.2% 1-year and 10-year productivity impacts respectively reported for management interventions in post-war Italy reported in Table 3 of Giorcelli (2017).[16]

We note that estimated productivity (looms per employee) rose a statistically significant amount by 2017 and by a positive but non-significant amount between 2011 and 2014, while the set of practices in the treatment plants was contracting. We can offer two possible explanations. First, the practices dropped actually may have had negative value. Second, we may be observing learning by doing, perhaps ignited by the changes in processes brought in by the consultants that opened new opportunities for learning or by unobserved management changes that encouraged learning.

In column (4) we asked the plants if they had used any consultants since 2011, and if so for how many days. Many of these firms had, and indeed, as column (5) shows, this use of consultants was significantly higher in the treatment plants. These consultants were local firms offering very practical advice on loom-changing practices, fabrics, human resources, or textile marketing, rather than the types of expensive international-firm management consulting provided by our intervention. We interpret this as a revealed preference indicator that treatment firms found

[15] We address the concern that outliers may be driving the results by winsorizing the top and bottom 10% of the data (each year) and find that the results do not change substantively.

[16] The results are also similar to the 1-year impact of 17% reported in Bloom et al. (2013).

the intervention useful and were more willing to pay for commercial consulting in the future. This was more likely to occur once some time had passed since their previous consulting experience in our project (panel B).

Finally, in column (5) we look at the adoption of marketing practices. Marketing practices were not part of our initial intervention, and so this enables us to examine whether changes in the specific practices on which our intervention focused are accompanied by broader management changes. Our measure is a score given for the adoption of seven practices: (1) does a director regularly attend trade shows; what is the frequency of systematically analyzing markets, products and prices to assess policies (and make changes wherever necessary) ((2), (3) and (4)); (5) does the firm have a dedicated brand; (6) does the firm have a sales and marketing professional; and (7) does the firm use any e-commerce (for sales) and social media (for advertising). Panel A shows that treatment firms are significantly more likely to adopt these marketing practices. Discussions with firms highlighted their attempts to be more systematic in management across a range of activities. So, in this sense, there were cross-practice management spillovers. Improving production and human-resource management practices led firms to value a more data-driven, systematic management approach, and hence apply this to other areas like marketing.

IV. CONCLUSIONS

In summary, the intervention in 2008-2010 did have lasting effects, but not the multiplier effect of on-going further improvements that the "Toyota Way" theory would have predicted. Indeed, a significant fraction of the induced improvements were dropped, especially if the plant manager changed, the directors were short of time, or if the practices were not common before the intervention. Still, many of the changes persisted, and spread throughout the treatment firms. There was also some persistence and some drop in the control plants' set of practices. Thus, the "inappropriate technologies" view does not find much support. Beyond that, the "three-year life" conventional wisdom described in the introduction is also decisively rejected, at least for the sort of practice changes our intervention induced.

The treatment firms were still much better managed in 2017 than the control, and key practices around quality control and inventory management were maintained. Moreover, the treatment firms used more consulting and did more marketing, suggesting that the more systematic approach to

management introduced by the intervention was spreading to areas the intervention had not addressed, and we see long-term benefits in terms of a measure of worker productivity. These lasting impacts highlight the importance of management in explaining persistent productivity differences amongst firms. Understanding why more firms do not invest in improving management, and what types of policies can change this, is therefore an important question for future research.

References

Anderson, Steven, Rajesh Chandy and Bilal Zia "Pathways to Profits: Identifying Separate Channels of Small Firm Growth Through Business Training", *Management Science*, forthcoming.

Bandiera, Oriana, Renata Lemos, Andrea Prat and Raffaella Sadun, (2017) "Managing the Family Firm: Evidence from CEOs at Work.", *Review of Financial Studies*, forthcoming.

Bennesden, Morten, Kasper Nielsen, Francisco Pérez-Gonzáles and Daniel Wolfenzon, (2007). "Inside the Family Firm: The Role of Families in Succession Decisions and Performance", *Quarterly Journal of Economics*, 122(2), 647-691.

Bertrand, Marianne and Antoinette Schoar, (2003). "Managing with Style: the Effect of Managers on Firm Policies," *Quarterly Journal of Economics*, 118(4), 1169–1208.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie and John Roberts (2013) "Does Management Matter? Evidence from India", *Quarterly Journal of Economics*, 128(1): 1-51

Bloom, Nicholas and Van Reenen, John (2007), "Measuring and Explaining Management Practices across Firms and Countries", *Quarterly Journal of Economics.* 122(4), 1351-1408

Braguinsky, Serguey, Atsushis Ohyama, Tetsuji Okazaki and Chad Syverson, (2015). "Acquisition, Productivity and Profitability: Evidence from the Japanese Cotton Spinning Industry." *American Economic Review,* 105(7): 2086-2119.

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar (2018), "The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico" *Journal of Political Economy*, 126(2), 635-687.

Capelli, Peter and David Neumark, (2001). 'Do 'High-Performance' Work Practices Improve Establishment-Level Outcomes?', *Industrial and Labor Relations Review*, 54(4): 737-775.

Clark, Greg (1987). "Why Isn't the Whole World Developed? Lessons from the Cotton Mills" *Journal of Economic History*, 47(1), 141-173.

Dellavigna,Stefano and Devin Pope (2017). "Predicting Experimental Results: Who Knows What?", *Journal of Political Economy*, forthcoming.

The Economist (2009). "Good to great to gone", July 7.

Fryer, Roland (2017). "Management and Student Achievement: Evidence from a Randomized Field Experiment", Harvard Working Paper.

Giorcelli, Michela (2017). "The Long-Term Effects of Management and Technology Transfer: Evidence from the US Productivity Program", UCLA Mimeo.

Groh, Matthew, Nandini Krishnan, David McKenzie and Tara Vishwanath (2016). "The Impact of Soft Skills Training on Female Youth Employment: Evidence from a Randomized Experiment in Jordan", *IZA Journal of Labor and Development*, 5(9).

Higuchi, Yuki, Edwin Mhede, and Tetsushi Sonobe (2016). "Short- and Longer-Run Impacts of Management Training: An Experiment in Tanzania", Mimeo. National Graduate Institute for Policy Studies, Tokyo.

Hirschleifer, Sarojini, David McKenzie, Rita Almeida and Cristobal Ridao-Cano (2016). "The Impact of Vocational Training for the Unemployed: Experimental Evidence from Turkey", *Economic Journal*, 126(597), 2115-2146.

Hsieh, Chiang-Tai, and Pete Klenow (2010). "Development Accounting," *American Economic Journal: Macroeconomics,* 2(1), 207-223.

Huselid, Mark (1995). "The Impact of Human Resource Management Practices on Turnover, Productivity and Corporate Financial Performance", *Academy of Management Journal*, 38: 635-672.

Ichniowski, Casey, Kathryn L. Shaw, and Giovanna Prennushi, (1997). "The Effects of Human Resource Management Practices on Productivity," *American Economic Review*, 86(3), 291-313.

Karlan, Dean, Ryan Knight, and Christopher Udry (2015). "Consulting and Capital Experiments with Microenterprise Tailors in Ghana", *Journal of Economic Behavior and Organization,* 118, 281-302.

Kiechel, Walter (2012). "The Management Century", *Harvard Business Review,* November.

Lazear, Edward, Kathryn Shaw and Christopher Stanton (2015). "The Value of Bosses", *Journal of Labor Economics,* 33(4), 823-61.

Liker, Jeffrey K. (2004*). The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer.* McGraw-Hill

Marshall, Alfred, (1887), "The Theory of Business Profits", *Quarterly Journal of Economics*, 1(4), 477-481.

McKenzie, David, and Christopher Woodruff (2014). "What Are We Learning from Business Training Evaluations around the Developing World?", *World Bank Research Observer*, 29(1), 48-82.

Osterman, Paul, 1994. 'How Common Is Workplace Transformation and Who Adopts It?', *Industrial and Labor Relations Review*, 47(2), 173-188.

Sirkin, Harold, Perry Keenan and Alan Jackson (2005). "The Hard Side of Change Management", *Harvard Business Review*, October.

Roberts, John (2018), "Needed: More Economic Analyses of Management", *International Journal of the Economics of Business,* forthcoming.

Syverson, Chad. 2011. "What Determines Productivity?", *Journal of Economic Literature,* 49(2), 326-365.

Walker, Francis (1887), "On the Sources of Business Profits", *Quarterly Journal of Economics,*1(3), 265-288.

Appendix

AI) Plant sample:

Table A2 reports the sample of plants by the four types (treatment and control, experimental and non-experimental). As noted in the text, one treatment firm exited because of the death of the owner without any male heirs, which led to the closure of one plant. Two more treatment plants closed because they were amalgamated into other plants within the same firm – that is, all the looms and equipment were moved onto one site for production economies of scale. We count these as a plant closure (since that plant stopped operating) but the output of that plant will still be included at the firm-level. Finally, both treatment and control firms opened some plants over this period due to demand growth.

AII) Management survey in 2011 and Imputing Labor Productivity:

Between November 2011 and January 2012 we ran an in-person survey of textile firms around Mumbai with 100 to 1,000 employees, using the Ministry of Commercial Affairs registry of firms plus a combination of industry lists, internet searches, and referrals as a sample frame (see online Appendix A2 of Bloom et al, 2013 for more sampling details). We identified 172 such firms, and were able to interview 113 of them (17 project firms and 96 non-project firms). The main purpose of this survey was to benchmark the management practices of our experimental sample against the industry as a whole, and we found that our project firms did not differ significantly in management practices from the non-project firms interviewed.

The interview followed a relatively standardized script, asking background questions about the firm (age, ownership, family involvement, markets etc), followed by questions about plant size (employees, output, plant numbers, production quantity), management practices, organizational structure, computerization, prior consulting, prior knowledge of the Stanford-World Bank project (we skipped this question for firms involved in the experiment), and any potential interest in future consulting waves. The full survey is available at www.stanford.edu/~nbloom/Template.xlsx.

In this paper, we use the data collected in this survey on the annual physical output of the firm (in meters or production picks), the number of employees (permanent plus contract), and the number of looms in the firm. We attempted to collect this for four years 2008-2011, and we were able to collect this information for all four years for 87 firms, and for two or three years for a further 7 firms. Using this data, we construct labor productivity as the log of physical production units per worker. This is similar to the sales per worker term often using to measure labor productivity, but has the advantage of not incorporating price effects.

Appendix Figure A2 shows the strong correlation (0.561) between labor productivity and looms per employee. Appendix Table A4 presents the corresponding regression relationship. Column 1 shows the strong cross-sectional relationship, which persists after adding year fixed effects (column 2), firm fixed effects (column 3), and both year and firm fixed effects (column 4). Column 4 then shows that annual changes in looms per employee are associated with changes in labor productivity. This yields the fitted relationship:

*Log production per worker = 0.734 (s.e. 0.114) * Log looms per worker + year effect + firm fixed effect.*

We use this fitted relationship to impute labor productivity impacts from our impact on looms per worker in Table 5.

Table A1: The textile management practices adoption rates

Area		Specific Practice	2008	2011	2017
Factory Operations	1	Preventive maintenance is carried out for the machines	0.4	0.7	0.95
	2	Preventive maintenance is carried out per manufacturer's recommendations	0.1	0.15	0.15
	3	The shop floor is marked clearly for where each machine should be	0.1	0.3	0.25
	4	The shop floor is clear of waste and obstacles	0.05	0.3	0.3
	5	Machine downtime is recorded	0.6	0.9	0.9
	6	Machine downtime reasons are monitored daily	0.45	0.9	0.85
	7	Machine downtime analyzed at least fortnightly & action plans implemented to try to reduce this	0..05	0.65	0.6
	8	Daily meetings take place that discuss efficiency with the production team	0.05	0.7	0.8
	9	Written procedures for warping, drawing, weaving & beam gaiting are displayed	0.1	0.45	0
	10	Visual aids display daily efficiency loomwise and weaverwise	0.25	0.7	0.4
	11	These visual aids are updated on a daily basis	0.15	0.6	0.25
	12	Spares stored in a systematic basis (labeling and demarked locations)	0.1	0.2	0.4
	13	Spares purchases and consumption are recorded and monitored	0.5	0.55	0.35
	14	Scientific methods are used to define inventory norms for spares	0	0.05	0.1
Quality Control	15	Quality defects are recorded	0.95	1	1
	16	Quality defects are recorded defect wise	0.25	0.85	0.95
	17	Quality defects are monitored on a daily basis	0.3	1	0.5
	18	There is an analysis and action plan based on defects data	0.05	0.7	0.3
	19	There is a fabric gradation system	0.55	0.85	1
	20	The gradation system is well defined	0.45	0.85	0.45
	21	Daily meetings take place that discuss defects and gradation	0.15	0.75	0.3
	22	Standard operating procedures are displayed for quality supervisors & checkers	0.05	0.6	0
Inventory Control	23	Yarn transactions (receipt, issues, returns) are recorded daily	0.89	1	1
	24	The closing stock is monitored at least weekly	0.28	0.83	0.56
	25	Scientific methods are used to define inventory norms for yarn	0	0	0
	26	There is a process for monitoring the aging of yarn stock	0.28	0.538	0.72
	27	There is a system for using and disposing of old stock	0.05	0.78	0.56
	28	There is location wise entry maintained for yarn storage	0.28	0.61	0.5
Loom Planning	29	Advance loom planning is undertaken	0.35	0.55	0.1
	30	There is a regular meeting between sales and operational management	0.5	0.6	0.45
Human Resources	31	There is a reward system for non-managerial staff based on performance	0.6	0.7	0.6
	32	There is a reward system for managerial staff based on performance	0.3	0.45	0.2
	33	There is a reward system for non-managerial staff based on attendance	0.35	0.5	0.5
	34	Top performers among factory staff are publicly identified each month	0.15	0.25	0.2
	35	Roles & responsibilities are displayed for managers and supervisors	0.05	0.5	0.5
Sales and Orders	36	Customers are segmented for order prioritization	0	0	0.11
	37	Orderwise production planning is undertaken	0.67	0.89	1
	38	Historical efficiency data is analyzed for business decisions regarding designs	0	0.1	0.08
All		Average of all practices	0.271	0.576	0.466

Notes: Reports the 38 individual management practices for all treatment plants (both experimental and non-experimental, unbalanced panel) in 2008, 2011 and 2017.

Table A2: Plant count

	2008	2011	2014	2017
Treatment – experimental	14	14	11	11
Treatment – non-experimental	6	9	9	9
Control – experimental	6	6	6	6
Control – non-experimental	2	2	4	4
Total	28	31	30	30

Notes: Lists the total number of plants in 2008 to 2017, including all dead and alive plants. One firm closed in 2014, so the total number of firms was 17, 17, 16 and 16 across the first four columns.

Table A3: Practice stickiness

		Adopted	**Dropped**	**Share Dropped**
9	Written procedures for warping, drawing, weaving & beam gaiting are displayed	7	7	1.00
22	Standard operating procedures are displayed for quality supervisors & checkers	11	10	0.91
11	These visual aids are updated on a daily basis	11	7	0.64
10	Visual aids display daily efficiency loomwise and weaverwise	11	6	0.55
21	Daily meetings take place that discuss defects and gradation	13	7	0.54
18	There is an analysis and action plan based on defects data	14	7	0.50
17	Quality defects are monitored on a daily basis	16	6	0.38
4	The shop floor is clear of waste and obstacles	6	2	0.33
33	There is a reward system for non-managerial staff based on attendance	9	3	0.33
20	The gradation system is well defined	8	2	0.25
24	The closing stock is monitored at least weekly	13	3	0.23
7	Machine downtime analyzed at least fortnightly & action plans implemented to try to reduce this	15	3	0.20
8	Daily meetings take place that discuss efficiency with the production team	19	3	0.16
5	Machine downtime is recorded	9	1	0.11
6	Machine downtime reasons are monitored daily	13	1	0.08
27	There is a system for using and disposing of old stock	15	1	0.07
1	Preventive maintenance is carried out for the machines	10	0	0.00
12	Spares stored in a systematic basis (labeling and demarked locations)	6	0	0.00
16	Quality defects are recorded defect wise	20	0	0.00
19	There is a fabric gradation system	9	0	0.00
26	There is a process for monitoring the aging of yarn stock	11	0	0.00
28	There is location wise entry maintained for yarn storage	7	0	0.00
35	Roles & responsibilities are displayed for managers and supervisors	9	0	0.00
37	Orderwise production planning is undertaken	6	0	0.00

Notes: Lists the practices ordered by the share of adopters between 2008 and 2011 that subsequently dropped them by 2017.

Table A4: Looms per employee and labor productivity

Dependent variable: Log(output/employees)	(1)	(2)	(3)	(4)
Log(looms/employee)	0.698	0.698	0.736	0.734
	(0.138)	(0.139)	(0.113)	(0.114)
Year fixed effects	No	Yes	No	Yes
Firm fixed effects	No	No	Yes	Yes
Firms	94	94	94	94
Observations	366	366	366	366

Notes: Regression results from the 2011 survey (detailed in Appendix A2). Only firms with non-zero and non-missing production picks, looms and employment are included. The dependent variable is production picks per employee (in logs). Regressions clustered at the firm level.

Table 1: The field experiment sample pre-intervention (2008)

	All				Treatment	Control	Diff
	Mean	Median	Min	Max	Mean	Mean	p-value
Sample sizes:							
Number of plants	28	n/a	n/a	n/a	19	9	n/a
Number of experimental plants	20	n/a	n/a	n/a	14	6	n/a
Number of firms	17	n/a	n/a	n/a	11	6	n/a
Plants per firm	1.65	2	1	4	1.73	1.5	0.393
Firm/plant sizes:							
Employees per firm	273	250	70	500	291	236	0.454
Employees, experimental plants	134	132	60	250	144	114	0.161
Hierarchical levels	4.4	4	3	7	4.4	4.4	0.935
Annual sales $m per firm	7.45	6	1.4	15.6	7.06	8.37	0.598
Current assets $m per firm	8.50	5.21	1.89	29.33	8.83	7.96	0.837
Daily meters, experimental plants	5560	5130	2260	13000	5,757	5,091	0.602
Management and plant ages:							
BVR Management score	2.60	2.61	1.89	3.28	2.50	2.75	0.203
Management adoption rates	0.262	0.257	0.079	0.553	0.255	0.288	0.575
Age, experimental plant (years)	19.4	16.5	2	46	20.5	16.8	0.662

Notes: Data provided at the plant and/or firm level depending on availability. **Number of plants** is the total number of textile plants per firm including the non-experimental plants. **Number of experimental plants** is the total number of treatment and control plants. **Number of firms** is the number of treatment and control firms. **Plants per firm** reports the total number of other textiles plants per firm. Several of these firms have other businesses – for example retail units and real-estate arms – which are not included in any of the figures here. **Employees per firm** reports the number of employees across all the textile production plants, the corporate headquarters and sales office. **Employees per experiment plant** reports the number of employees in the experiment plants. **Hierarchical levels** displays the number of reporting levels in the experimental plants – for example a firm with workers reporting to foreman, foreman to operations manager, operations manager to the general manager and general manager to the managing director would have 4 hierarchical levels. **BVR Management score** is the Bloom and Van Reenen (2007) management score for the experiment plants. **Management adoption rates** are the adoption rates of the management practices listed in Table A1 in the experimental plants. **Annual sales ($m)** and **Current assets ($m)** are both in 2009 US $million values, exchanged at 50 rupees = 1 US Dollar. **Daily meters, experimental plants** reports the daily meters of fabric woven in the experiment plants. Note that about 3.5 meters is required for a full suit with jacket and trousers, so the mean plant produces enough for about 1600 suits daily. **Age of experimental plant (years)** reports the age of the plant for the experimental plants.

Table 2: Short and long run impact on management practices

Dep Var: Proportion of management practices implemented	(1)	(2)
Treatment*Year=2011	0.206***	0.249***
	(0.042)	(0.038)
	[0.003]	[0.001]
Treatment*Year=2017	0.197**	0.218**
	(0.062)	(0.057)
	[0.007]	[0.003]
Year=2017	-0.122***	-0.122***
	(0.016)	(0.016)
	[0.732]	[0.694]
Baseline 2008 Management Score	0.668**	0.878***
	(0.219)	(0.176)
	[0.022]	[0.006]
P-value of test of equality of treatment in 2011 and 2017	0.802	0.457
Sample Size	37	34

Notes: Notes: Robust standard errors in () parentheses and permutation test p-values in [] parentheses. Both are clustered at the firm level. *, **, and *** denote significance at the 10, 5, and 1 percent levels respectively on the robust standard errors. Permutation tests report the p-value for testing the null hypothesis that the treatment had no effect by constructing the permutation distribution of the estimator using 4000 possible permutation of firm-level random assignment. The second column limits the sample from column 1 to plants that were present in both years with no missing management scores.

Table 3: Reasons for the change in management practices

	Treatment Experimental	Treatment Non-Experimental	Control Experimental	Control Non-Experimental	All
Added Practices (%)					
New manager	1.2	0.6	0.4	0	0.8
Product, customer or equipment change	0.7	1.8	0	0	0.9
Spillovers from other firms	0.7	0.3	2.2	2.7	1.1
Spillovers from other plants in the same firm	0	4.2	0	0	1.3
Total	2.6	6.9	2.6	2.7	4.1
Dropped Practices (%)					
New Manager	9.9	0.6	1.8	1.4	4.6
	2.9	3.0	5.3	1.4	4.2
Reduced directors time	3.9	3.0	3.6	4.1	3.6
Total	16.7	6.6	10.7	6.9	12.4
No Change (%)	80.7	86.4	86.7	90.4	83.5
Total	100	100	100	100	100

Notes: Lists the shares of practice by plant cells in terms of reasons for change between 2011 and 2017 in terms of practices added, dropped or left unchanged. Calculated as a share of 1,042 practices, which are comprised of the 38 practices across the 28 plants (11 treatment experimental, 9 treatment non-experimental, 6 control experimental and 2 control non-experimental) in operation in both 2011 and 2017, except for the inventory practices which are missing in plants which hold no inventory because they make to order.

Table 4: Determinants of changes in management from 2011 to 2017

DV=0/1/-1 management score change	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Experimental plant		-0.128**				-0.098***	-0.097***
		(0.046)				(0.021)	(0.022)
Treatment plant			0.020			0.047	0.043
			(0.037)			(0.029)	(0.023)
New plant manager*treated				-0.103**		-0.096**	-0.075*
				(0.047)		(0.038)	(0.045)
New plant manager*control				-0.035		-0.007	-0.010
				(0.029)		(0.027)	(0.036)
Frequency of practice usage in 2008					0.095**	0.095**	0.095**
					(0.037)	(0.037)	(0.037)
Management score in 2011							-0.132
							(0.160)
Constant	-0.083***	0.050	-0.101***	-0.048**	-0.111***	-0.052*	-0.052*
	(0.027)	(0.046)	(0.015)	(0.023)	(0.028)	(0.027)	(0.027)
Observations	1,042	1,042	1,042	1,042	1,042	1,042	1,042

Notes: Dependent variable is the change in the -1,0,1 indicator for the change in management practice between 2011 and 2017. The sample is the 38 practices across the 28 plants (11 treatment experimental, 9 treatment non-experimental, 6 control experimental and 2 control non-experimental) in operation across both periods, except for the inventory practices which are missing in plants which hold no inventory because they make to order. Regressions clustered at the firm level. *** denotes 1%, ** denotes 5%, * denotes 10%

Table 5: Longer-run Firm performance and management changes

Dep Var	Looms (in logs) (1)	Employees (in logs) (2)	Looms per employee (in logs) (3)	Consulting days (in logs) (4)	Marketing practices (score) (5)
Panel A: Long-run performance					
Treatment_i*(Year>=2011)$_t$	-0.032	-0.269	0.237**	1.324**	1.361**
	(0.226)	(0.277)	(0.090)	(0.556)	(0.618)
	[0.86]	[0.27]	[0.030]	[0.103]	[0.068]
Panel B: Treatment impact by period					
Treatment_i*(Year==2011)$_t$	-0.041	-0.141	0.100	0.000	1.197**
	(0.213)	(0.269)	(0.115)	(0.000)	(0.528)
	[0.837]	[0.625]	[0.446]	[1.00]	[0.105]
Treatment_i*(Year==2014)$_t$	-0.204	-0.413	0.209	1.576*	-0.068
	(0.253)	(0.333)	(0.120)	(0.859)	(0.074)
	[0.360]	[0.168]	[0.156]	[0.252]	[0.212]
Treatment_i*(Year==2017)$_t$	0.149	-0.263	0.412***	2.491**	2.965*
	(0.302)	(0.298)	(0.138)	(1.040)	(1.469)
	[0.585]	[0.337]	[0.004]	[0.098]	[0.068]
p-value for F-test Treatment_i*(Year==2011) & Treatment_i*(Year==2014)$_t$ & Treatment_i*(Year==2017)$_t$	0.036	0.177	0.230	0.083	0.088
Control group mean	4.271	5.021	-0.750	0.067	0.583
Years	2008, 11, 14, 17	2008, 11, 14, 17	2008, 11, 14, 17	2008, 11, 14, 17	2008, 11, 14, 17
Firms	17	17	17	17	17
Observations	66	66	66	66	66

Notes: Data for pre-treatment (2008) and post-treatment (2011, 2014 and 2017) years, except firms for which basic performance data was missing. Sales and marketing practices is an indicator from 0 to 10 defined as the count of ten 0/1 Sales and Marketing practices like "Attending trade shows", "Hiring sales and marketing professionals", "Analyzing product portfolios", "Setting up a firm brand". Regressions clustered at the firm level and standard errors in parentheses. *** denotes 1%, ** denotes 5%, * denotes 10%. F-test reports p-value of the joint test testing the equality of the treatment effects over all three post-treatment periods. Permutation tests in [ ] below report the p-value for testing the null hypothesis that the treatment had no effect by constructing the permutation distribution of the estimator using 4000 possible permutation of firm-level random assignment.