Review of Economic Studies (2020) **0**, 1–8 doi:10.1093/restud/rdaa002

Managerial Quality and Productivity Dynamics

ACHYUTA ADHVARYU

University of Michigan, Good Business Lab, NBER, BREAD, William Davidson Institute.
E-mail: adhvaryu@umich.edu

ANANT NYSHADHAM

University of Michigan, Good Business Lab, NBER, BREAD.
E-mail: nyshadha@umich.edu

JORGE TAMAYO

Harvard University, Harvard Business School.
E-mail: jtamayo@hbs.edu

First version received July 2019; *Accepted August* 2021 (*Eds.*)

Do productivity and managerial quality vary within the firm? If so which managerial traits and practices matter most for team productivity? Combining granular garment production data with survey data on managers across 120 production lines in India, we document substantial productivity dispersion both across teams producing overlapping products and within team over the course of production runs, and structurally link this variation to a comprehensive assessment of supervisor quality. We find that factors related to managerial attention and control are the most important for enabling line productivity, both more impactful than traditionally emphasized dimensions like cognitive skills and tenure. We document that one mechanism by which specific managerial practices contribute to productivity is by way of enabling faster learning-by-doing. In-sample pay patterns suggest potential net gains from screening for or training in less readily measured dimensions of managerial quality, as pass-through of productivity contributions to pay is incomplete.

Keywords: management, productivity, non-cognitive skills, learning-by-doing, ready-made garments, India.
JEL Codes: M50, J24, J33, L2

1. INTRODUCTION

A long and influential empirical literature has documented a strong relationship between managerial quality and productivity using two parallel approaches. The first has used rich survey data on a broad array of management characteristics and practices of establishments and top-level executives to explain productivity variation across firms even within the same industries and country contexts (Bandiera et al., 2020; Bloom et al., 2016; Bloom and Van Reenen, 2007, 2011), but is limited in its ability to separate establishment level features like organizational structure

The editor in charge of this paper was Uta Schoenberg.

and production technology from the quality of the manger. The second has used "insider" firm data to show that managers vary widely in the productivity of subordinates even within the firm (Bertrand and Schoar, 2003; Lazear et al., 2015), but stops short of unpacking which specific characteristics and practices of managers contribute most to productivity among subordinates.

In this study, we combine these two approaches to build on recent "insider" work identifying some specific ways in which managers can enable productivity of workers and teams within the firm, for example, through retention (Hoffman and Tadelis, 2018), elicitation of effort (Frederiksen et al., 2020), and task allocation (Adhvaryu et al., 2019). We match granular productivity data from 120 production lines inside one of the largest garment export firms in the world with rich survey data on a comprehensive array of skills, traits, and practices of mid-level managers to identify which dimensions of managerial quality matter most for productivity when holding organizational structure and production technology fixed. We begin by documenting substantial variation in productivity both across nearly identical production teams producing overlapping products as well as within teams across time as experience evolves over the course of a production run.

We then show that managers vary in quality along a myriad of dimensions and that this dispersion in managerial quality coincides with variation in productivity across the lines they supervise. The relationship between the quality of the supervisor and the productivity of the team appears nuanced and complex in that some dimensions of quality map to higher minimum or initial productivity and others to higher maximum or peak productivity or both. Accounting for garment style assignment across lines as well as worker composition, we leverage the high degrees of freedom in our matched production and survey data to estimate a structural model of how productivity varies both within a line across the life of an order as well as across lines with supervisors whose qualities differ along multiple, potentially interactive, dimensions.

Our empirical strategy proceeds in three steps. First, we estimate a canonical learning function, taking a form similar to the functions estimated in, e.g., Benkard (2000) and Levitt et al. (2013), except that we allow for the parameters governing the shape of the learning curve to vary by managers. Second, in the spirit of Attanasio et al. (2015); Cunha et al. (2010), we estimate a nonlinear latent factor model using the data from our managerial survey to recover information about the joint distribution of k latent factors of managerial quality and the learning parameters estimated in the first stage. The data reveal seven distinct factors falling into three broad categories: *commonly screened characteristics* (Tenure and Demographics), *less readily screened characteristics* (Cognitive Skills, Control, and Personality), and *trainable practices and behaviors* (Autonomy and Attention). We allow for the recovered factors to be correlated with each other and to interact nonlinearly in the determination of productivity. Finally, we draw a synthetic dataset from the joint distribution of these factors and the productivity parameters and estimate a CES-type function for each learning parameter with the factors of managerial quality as arguments.

We find overall that the two most important dimensions of managerial quality for enabling line productivity appear to be Attention, measured by monitoring frequency and effort invested in personnel management, and Control, reflecting a belief in one's ability to affect change rather than acquiesce to chance or predetermination. Both of these dimensions are more impactful than traditionally emphasized dimensions like Cognitive Skills and Tenure, but have received relatively little emphasis in the literature to date. A few recent studies have begun to model and document the impacts of managerial inattention on firm performance (Bandiera et al., 2014; Halac and Prat, 2016; Hortaçsu et al., 2017), but our results are among the first to emphasize the importance of internal locus of control.

The structure imposed in our approach allows us to glean additional insights from counterfactual simulations of screening and training policies among manufacturers specializing in different scales of production. Many small scale domestic producers in India and other similar developing countries with large garment industries indeed specialize in smaller orders; while export suppliers to multinational brands tend to produce disproportionately large volume orders. Simulations reveal that Cognitive Skills, Control, and, by way of correlations with other factors, Demographics should seemingly be emphasized irrespective of the length of the order; while Tenure, Autonomy, Attention and, again due to correlations with other factors, Personality should be of particular interest to factories producing primarily large volume orders.

Finally, we find that in-sample pay reflects imperfect pass-through of the productive value of managerial quality to supervisor pay. Specifically, simulations show that the firm could perhaps benefit most from screening on Control and training in Attention without likely having to increase pay as commensurately. This represents an opportunity to improve managerial quality and resulting productivity at low incremental wage cost, at least in our context of labor-intensive manufacturing in India.

This pattern is consistent with evidence across many country and industry contexts that the skills, traits, and practices that enable managers to lead effectively are often hard to observe and measure (Burks et al., 2015; Dustmann et al., 2015; Hoffman et al., 2017; Kahn, 2013; Kahn and Lange, 2014; Lange, 2007; Schönberg, 2007). Recent studies on training and signaling interventions in low-income country labor markets have emphasized similar frictions (Adhvaryu et al., 2018; Alfonsi et al., 2017; Bassi and Nansamba, 2017), and a related study from the Indian manufacturing context found a limited role for performance pay among managers (Bloom et al., 2013). Additionally, taken together the results of these simulations are remarkably consistent with findings from a randomized training trial in similar factories in which managerial skills training for production line supervisors generated gains in their Attention and Autonomy and resulting improvements in the productivity of lines they supervise, but negligible effects on their pay (Adhvaryu et al., 2021).

We contribute to the insider literature on productivity dispersion across subordinates of managers within the firm (Bertrand and Schoar, 2003; Lazear et al., 2015). We add to recent advances on specific ways in which managers enable their subordinate workers (Adhvaryu et al., 2019; Frederiksen et al., 2020; Hoffman and Tadelis, 2018) by leveraging rich primary survey data to compare contributions to productivity across a broad array of managerial skills, traits and practices. Our results reiterate previous findings on the importance of managerial attention (Bandiera et al., 2014; Ellison et al., 2018; Halac and Prat, 2016; Hortaçsu et al., 2017) in addition to traditional dimensions of quality like tenure and cognitive skills. Our results also emphasize a novel dimension of quality in Control, reflecting a belief in the ability to affect change rather than acquiesce to chance or predetermination. The importance of this dimension far outweighs that of oft-cited personality measures of conscientious and perseverance (Almlund et al., 2011; Borghans et al., 2008; Deming, 2017; Donato et al., 2017) as well as demographics and measures of discrimination (Bertrand and Duflo, 2017; Hjort, 2014).

We also contribute to the literature on management and productivity (Bandiera et al., 2020; Bloom et al., 2013, 2016; Bloom and Van Reenen, 2007), by leveraging within firm variation in rich survey measures of comprehensive manager quality and productivity dispersion both across teams and within teams across orders. Much of the existing work has studied managerial quality at the establishment or firm level and highlighted key practices such as monitoring, personnel management (components of our Attention factor), and decentralization (reflected in our Autonomy measure) (Aghion et al., 2017; Bloom et al., 2014; Bloom and Van Reenen, 2011). Our results show that these relationships between Attention or Autonomy and productivity at the establishment or firm level are mirrored in mid-level management of teams within the firm, even after accounting for personality traits and skills of the individual managers. Furthermore, we are able to document that one mechanism by which these specific managerial practices contribute to productivity, at least in manufacturing, is by way of enabling faster learning-by-doing on production lines, consistent with empirical evidence from other contexts such as automobile assembly (Levitt et al., 2013).

2. DATA

We use data from two main sources for this study. The first source is data on the style being produced and productivity achieved across lines and days, and the second is survey data on the characteristics and practices of the permanently assigned managers of the lines (Shahi Exports Pvt Ltd, 2018).

2.1. *Production Data*

We use line productivity data at the daily level for two years, from March 2013 to July 2016, from six garment factories in Bengaluru, India. The data include the style or product on which the line is working, the number of garments the line assembles, and the target quantity for each day. Target quantities are lower for more complex garments (since lines can produce fewer complex garments in a given day), and therefore are an appropriate way to normalize productivity across lines producing garments of varying complexity. Our primary measure of productivity is efficiency, which equals garments produced divided by the target quantity for that line-order-day. Efficiency is the global industry standard measure of productivity in garments.

The target quantity for a given garment is calculated using a measure of garment complexity called the standard allowable minute (SAM). SAM is taken from a standardized global database of garment industrial engineering that includes information on the universe of garment styles. It measures the number of minutes that a particular garment should take to produce. For instance, a line producing a style with SAM of 30 is expected to produce 2 garments per hour per worker on the line. Accordingly, a line of 60 workers producing a style with SAM of 30 for 8 hours in a day will have a daily target of 960 units.[1] If the line produces 600 garments by the end of the day its efficiency would be $600/960 = .625$ for that day. We use daily line-level efficiency as the key dependent variable of interest.[2]

Figure 1 plots the distribution of efficiency across line-day observations. We see that, though line-day efficiency has a mean of roughly 50%, there is a great deal of variation. A line may achieve as little as less than 10% efficiency on a day or nearly 100%.[3] In attempting to determine the drivers of this substantial variation, one naturally asks how much productivity varies systematically across lines, perhaps due to varying managerial quality, as compared to within line across days. The economics literature has placed considerable emphasis on learning dynamics in manufacturing productivity across various sectors (Benkard, 2000; Levitt et al., 2013; Thompson, 2012). Accordingly, we next investigate the importance of learning dynamics in our empirical context.

From the production data, we can calculate for how many days a production line has been producing a particular garment style and document how efficiency evolves over the life of a

1. That is, the line has 60 minutes $\times$ 8 hours $\times$ 60 workers $=$ 28,800 minutes to make garments that take 30 minutes each, so $28{,}800/30 = 960$ garments by the end of the day.

2. We have run all the same analysis with log quantity as the outcome instead of log efficiency and find qualitatively identical results (omitted for brevity but available upon request). We keep log efficiency as our preferred outcome as this most closely corresponds to outcomes used in related studies like defect rates in Levitt et al. (2013) and labor per unit produced Benkard (2000) and Thompson (2012).

3. We are told the rare instances in which a line produces nearly 0% efficiency most likely reflect batch-setting or machine calibration days.

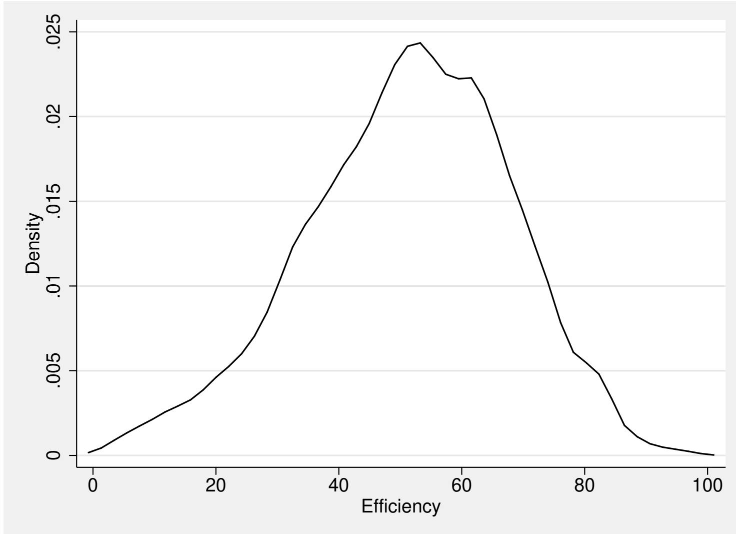

FIGURE 1
Dispersion in Line-day Productivity

Note: Figure 1 shows the distribution of the average efficiency across line-day observations, across the six factories included in our study. The data spans from March 2013 to July 2016. Our measure of productivity is daily efficiency, which equals the percentage of the target quantity of a particular garment that is achieved per day. The target quantity is calculated using a measure of garment complexity called the standard allowable minute (SAM), which is equal to the number of minutes that a particular garment should take to produce.

production run.[4] We see in Figure 2 that indeed learning contributes substantially to variation in productivity within lines across days. On average, a line begins each order at only 40% efficiency, but achieves peak efficiency of more than 55% after 3 weeks producing the same style.

We then split production lines into terciles on the basis of their average efficiency and repeat this exercise separately for each tercile. In Figure 3A, we see that productivity varies systematically across lines in addition to the variation in efficiency over the course of an order. Bottom tercile lines start orders at only 35% efficiency on average and peak around 50% efficiency; while top tercile lines start orders at 65% efficiency on average and peak at nearly 80%. That is, the systematic variation across lines is roughly twice the magnitude (i.e., roughly 30 percentage points) of the variation across the course of an order within line (i.e., roughly 15 percentage points).

The middle tercile in Figure 3A also draws attention to heterogeneity in learning dynamics across lines in addition to the level differences. That is, we see that these middle tercile lines have initial efficiency on orders closer to bottom tercile lines but increase their efficiency quickly and dramatically over the course of the order to close the substantial gap with top tercile lines. Figure 3B plots the differences between terciles to more clearly see this pattern. These large differences both in levels of efficiency and degree of learning across lines begs the question of what might be driving such substantial variation. Below we document large variation in a broad array of

4. We can measure learning-by-doing in 2 ways: as a function of the consecutive number of days that a line has been working on a particular style, or as a function of the cumulative quantity the line has produced of that style to date. By conducting our analysis of learning using a time-based measure of accrued experience, we circumvent the issue of endogenous productivity innovations across unit time. That is, serial correlation in production innovations are less concerning when the unit of experience is deterministic like time rather than stochastic like quantity produced to date. This issue is discussed and investigated in detail in previous studies. See, e.g., Thompson (2001).

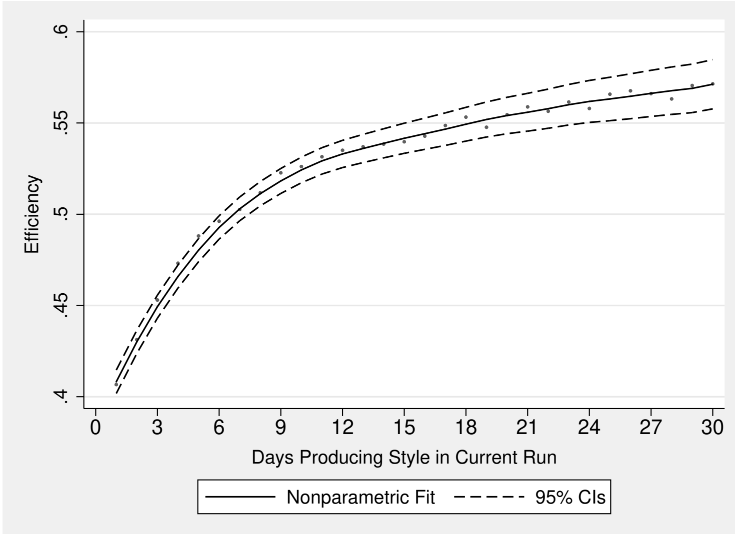

FIGURE 2

Learning (Efficiency by Days Running)

Note: Figure 2 depicts learning curves of efficiency by experience with experience defined by consecutive number of days a style has been running on the production line. The raw mean of efficiency by bin of experience is depicted in the scatter plot and the fitted curve (solid line) is the result of a lowess smoothed non-parametric estimation with a bandwidth of 0.4. Dashed lines represent 95% confidence intervals. Experience is trimmed at the 90th percentile in this graphical depiction to ignore outliers, but not from any regression analysis below.

managerial practices and characteristics and present preliminary evidence that this variation coincides with variation in efficiency.

We can also see in the data whether a line is producing a style that it has produced in the past, and how that changes current learning-by-doing. In particular, we define two variables that measure retained prior learning and forgetting: 1) the number of days since the production line last produced the style it is currently producing, and 2) the total number of days that the line produced the same style over prior production runs. Of course, these variables are positive only when lines have produced a particular style more than once and are all 0 when a line is running a style for the first time. We follow previous studies in combining these variables into an accumulated stock of experience, net of depreciation, for each style that each line is observed producing (Benkard, 2000; Levitt et al., 2013). We discuss this in greater detail in section 3 below.

Finally, we should also note that, though we use line-day productivity data in this study, the raw data from which these measures are constructed allow us to also observe which workers are working on each line each day. Related studies from nearly identical empirical contexts have documented variation in worker composition of lines across days and emphasized the importance of accounting for productivity effects of this variation (Adhvaryu et al., 2021,?). Accordingly, to account for any variation deriving from worker composition, we include worker fixed effects in our analysis.[5] To confirm the empirical feasibility of this, we report statistics documenting substantial worker mobility in Table A6 in the Appendix. The full set of worker fixed effects

5. Adhvaryu et al. (2021) show that including the full set of worker fixed effects is sufficient to purge line and manager level productivity parameters of variation deriving from worker composition. That is, they note that the production function in this context does not appear to exhibit manager-worker match effects.

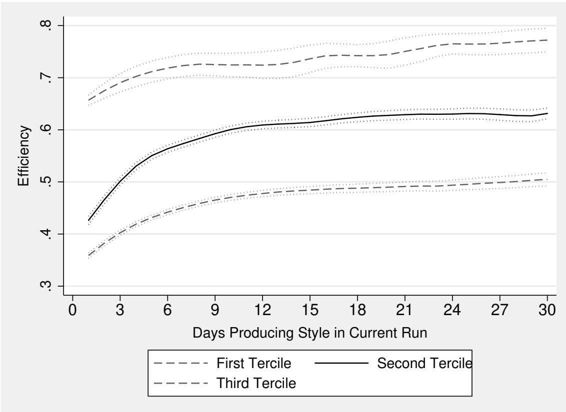

FIGURE 3A
Learning by Terciles

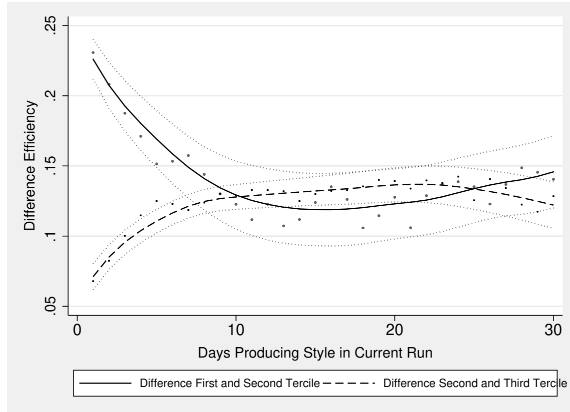

FIGURE 3B
Difference Between Terciles

Note: Figure 3A depicts learning curves of efficiency by experience with experience defined by consecutive number of days a style has been running on the production line, split into terciles on the basis of average efficiency. Figure 3B plots the difference between terciles. The raw mean of efficiency by bin of experience is depicted in the scatter plot in both figures and the fitted curve (solid line) is the result of a lowess smoothed non-parametric estimation with a bandwidth of 0.25. Dashed lines represent 95% confidence intervals. Experience is trimmed at the 90th percentile in this graphical depiction to ignore outliers, but not from any regression analysis below.

explains roughly 17.5% of the variation in line-day productivity, leaving a great deal of variation to be explained by managerial characteristics and learning.

Table 1 presents summary statistics of key variables of interest in the production data. We use data from 120 production lines with a total of 153 supervisors.[6] Our sample comprises nearly 50,000 production line-date observations, and we observe more than 2,700 line-style pairings. More than 40% of styles are observed being produced by more than one line. The median number of styles a line is observed producing over the course of the data is 27; and 88% of lines are observed producing the same style more than once. Mean efficiency is about 0.51 overall, but less than 0.41 on the first day of a new production run.

2.2. *Manager Pay*

In additional analysis, we explore the degree to which any contributions of various managerial quality measures to productivity translate into supervisor pay. Given the difficulty in accurately measuring dimensions of managerial quality, as outlined in our approach below, and the complexity and nuance in the relationships between dimensions of quality and various aspects of productivity, we might expect that the firm struggles to appropriately identify and reward supervisor quality. To investigate this, we obtained pay data for each supervisor from the month in which the survey was completed (November 2014).

These data include both monthly salary as well as any production bonus earned by the supervisor when the production line exceeded targets. Summary statistics for these pay variables

6. We restrict our analysis to the largest connected set of styles-lines, which contains 98.61% of the available data. We use the *bgl* toolbox in matlab to extract the largest connected set. Finally, we use an iterative conjugate gradient algorithm suggested by Abowd et al. (2002) to solve for the standard normal equations.

TABLE 1
Summary Statistics

	Observations	
Number of line-day observations	49,976	
Number of lines	120	
Number of styles	1,356	
Number of line-style matchings observed	2,742	
Number of supervisors	149	
Percent of lines producing same style more than once	88%	
Percent of styles produced at more than one line	41%	
Median of the number of different styles per line	27	
	Mean	**SD**
Production		
Efficiency	0.512	0.168
Initial Efficiency (first day of production run)	0.407	0.207
Current Experience		
Total length of production run in days	12.475	13.916
Experience from Prior Production Runs		
Total days of prior experience on a given style	19.316	24.687
Intervening days between runs of the same style	14.896	23.425

Note: Efficiency is equal to the garments produced divided by the target quantity of that particular garment. The target quantity is calculated using a measure of garment complexity called the standard allowable minute (SAM), which is equal to the number of minutes that a particular garment should take to produce.

are reported in the bottom rows of Table 2. Note that there appears only a negligible difference between the monthly salary alone and complete pay inclusive of production bonus. That is, while supervisors can in theory be rewarded for their productivity by way of production bonuses, these bonuses make up only a small fraction of supervisor compensation.[7] Accordingly, in order to appropriately reward supervisor quality in practice, the firm must adjust monthly salary to reflect quality. We explore the degree to which we observe this occurring below.

2.3. *Management Survey Data*

Each line is managed by 1 to 3 permanently assigned supervisors who allocate workers to tasks and are charged with motivating workers and diagnosing and solving production problems (such as machine misalignment or productivity imbalances across the line) to prevent and relieve bottlenecks and keep production on schedule. To measure managerial quality, we conducted a survey of all line supervisors. We drew from several sources to construct the management questionnaire, in particular borrowing heavily from Lazear et al. (2015), Schoar (2014), Bloom and Van Reenen (2010) and Bloom and Van Reenen (2011). The survey consisted of several different modules intended to measure both traditional dimensions of managerial skill like job and

7. This pattern is consistent with evidence from a previous study from a similar context (Bloom et al., 2013).

industry-specific tenure and cognitive skills as well as leadership style and specific managerial practices that have been emphasized in the literature. Additional modules on personality and risk and time preferences were also administered. Overall the survey covered work history, leadership style, management practices, personality psychometrics, cognitive skills, demographic characteristics and discriminatory attitudes.

In order to form a comprehensive assessment of each manager's "quality," we utilize the entirety of the survey in constructing measures to include in the non-linear factor system.[8] Following Cunha et al. (2010), Attanasio et al. (2015), and Attanasio et al. (2015), we allocate this full set of measures to factors by first conducting exploratory factor analyses within each module of the survey to determine if measures within a module appear to inform a single factor or multiple factors. We then pool measures across related modules (e.g., leadership style and managerial practices) and perform the exploratory factor analysis again on this pooled set to confirm that measures are being correctly mapped to the factor for which they are most informative.[9] This exercise, discussed in greater detail in Appendix B, statistically corroborates the intuitive and intentional mapping of measures into cohesive survey modules covering common managerial quality concepts. The resulting measures mapped to each factor are as follows:

- **Tenure**: The literature on productivity contributions of industry, firm, and job-specific accrued human capital, is large and well-established (Gibbons and Waldman, 2004; Jovanovic, 1979; Mincer and Ofek, 1982; Mincer et al., 1974; Neal, 1995; Topel, 1991). To capture Tenure, we use 4 measures: total years working, years working in the garment industry, years working as a garment line supervisor, and years supervising the current line.
- **Demographics**: Social preferences and demographic discrimination have also been shown to be determinants of team productivity (Hjort, 2014). We use two measures to capture demographic similarity between the supervisor and workers on the line they manage and any discriminatory attitudes the supervisor might have regarding demographic characteristics of their workers. The first is a simple count of the number of similarities between supervisor and majority of workers on the line in the following dimensions: age, gender, religion/caste, migrant status, and native language. The second measure is a count of the number of demographic dimensions (total of 9) over which the supervisor expressed no discriminatory preference.
- **Cognitive Skills**: The literature on returns to cognitive skills in productivity and earnings is nearly as long-standing and well-established as that for tenure (Boissiere et al., 1985; Bowles et al., 2001). To inform the Cognitive Skills factor, we use a measure of short-term memory and two measures of arithmetic skill. Digit span recall captures the largest number of digits in an expanding sequence the respondent was able to successfully recall. We use both the number of correct responses on a timed arithmetic test we administered as well as the percent of the attempted problems that had correct responses.
- **Personality**: Recent empirical studies have begun to document the incremental importance of personality psychometrics, alongside cognitive skills and specialized human capital accumulation, for earnings and productivity (Borghans et al., 2008; Heckman and Kautz, 2012). We included in the survey a standard module for conscientiousness meant to capture

8. In the end, we include all measures from the survey except for a few additional demographic (e.g., mode of transportation to work) and work history (e.g., second sources of income and agricultural experience) variables that were irrelevant to the research questions in this study.

9. Note that the measurement system we implement allows for the recovered factors to be correlated with each other, so it is permissible for measures to load incidentally onto other factors. However, we ultimately want to identify each factor from the set of measures which load primarily onto that factor. Accordingly, we check for each mapping that the measure most strongly informs the factor to which it is mapped above all other factors.

commonly measured personality psychometrics.[10] In addition, we collected measures of perseverance, self-esteem, and Kessler's psychological distress scale. The exploratory factor analysis described in Appendix B revealed that conscientiousness, perseverance, self-esteem, and psychological distress are highly correlated in the data and load onto a single factor which we call Personality.

- **Control**: We also collected survey measures of internal locus of control, risk aversion, and patience.[11] The internal locus of control module is meant to capture the degree to which a person believes in their own ability to control events or enact change as opposed to believing strongly in fate or pre-determination. The exploratory factor analysis described in Appendix B revealed that internal locus of control loads onto a distinct factor (which we call Control), apart from the above seemingly related personality traits. We also find that risk aversion and patience are highly correlated in the data with each other and load most strongly onto this same factor with internal locus of control.
- **Autonomy**: Finally, we collected survey measures of managerial behaviors and practices emphasized in previous studies. These modules measured leadership behaviors with respect to "initiating structure" and "consideration" (Stogdill and Coons, 1957) and specific management practices such as problem identification and solving and a self-assessments of managerial quality.[12] "Initiating structure" is said to capture the degree to which a manager plays a more active role in directing group activities; while "consideration" is meant to capture a good rapport with subordinates (Korman, 1966).[13] We find in the exploratory factor analysis described in Appendix B that both measures of leadership style ("initiating structure" and "consideration") load onto the same factor with initiating structure having the higher loading. Our two measures of the degree to which the supervisor takes the lead in and responsibility for identifying and solving production problems also load onto this same factor, along with a self-assessment measure of managerial quality relative to peers. Accordingly, we interpret this factor as capturing autonomy on the part of the supervisor, both in terms of leadership style and management practices. The empirical literature on the value of autonomy among lower level managers is small, but a few recent papers on decentralization of management have emphasized the importance of this dimension. Bloom and Van Reenen (2011) emphasize managerial autonomy/decentralization as an important dimension of managerial quality, drawing from earlier evidence of the value of autonomy at higher levels of organizational hierarchy (Groves et al., 1994). Aghion et al. (2017) find that more empowered lower-level management allows for stronger resilience during economic slowdowns. Bresnahan et al. (2002) find that the productivity returns to information technology are highest when management is decentralized. Relatedly, Acemoglu et al. (2007) find that firms closer to the technological frontier are more likely to decentralize, and Bloom et al. (2014) show that information technology enables firms to decentralize, in contrast to communication technology which promotes centralization.

10. Piloting showed that the other "Big Five" modules produced measures that were highly correlated with conscientiousness. This is consistent with what other recent studies have found among blue-collar workers in developing countries (Bassi and Nansamba, 2017). Accordingly, we did not administer the other Big 5 modules and rely on conscientiousness alone.

11. Modules for risk and time preferences were adapted from those used in the Indonesian Family Life Survey.

12. The module from which we obtain these measures is taken from the World Management Survey (Bloom and Van Reenen, 2007), adapted to allow for closed responses as opposed to open as piloting revealed closed response questions to be more effective in our setting with frontline supervisors in developing country factories.

13. These two behaviors are often hypothesized to be somewhat distinct from each other, but the factor analysis shows that in our context initiating structure and consideration are highly correlated. Nevertheless, both have been consistently validated as informative measures of successful leadership (Judge et al., 2004).

- **Attention**: The second factor from these management modules reflects contributions from five managerial practice measures: efforts to achieve production targets, production monitoring frequency, active personnel management, communication, and issues motivating workers and overcoming resistance. Each of these is meant to measure effort and attention on the part of the supervisor in accomplishing managerial tasks. The first measures the number of different practices the supervisor engages in to ensure production targets are met. The second records the number of times in a day the supervisor makes rounds of the production line to identify any production problems. The third measures the number of different practices the supervisor engages in to retain workers, motivate low performing workers, and encourage high performing workers. The fourth measures the frequency of communication regarding production with both workers and upper level managers, with a higher value representing less communication. The fifth measures the frequency with which the supervisor reports issues motivating workers and overcoming resistance to initiatives and change. Accordingly, we interpret this factor as capturing managerial attention. The literature on managerial attention is long-standing in theory and has added some recent empirical evidence (Ellison and Snyder, 2014; Reis, 2006). For example, Adhvaryu et al. (2019) find that more attentive managers are better able to diagnose and relieve bottlenecks that arise from shocks to worker productivity.

Summary statistics for these measures across all 153 supervisors are presented in Table 2.[14] Taken together, we see clear evidence of substantial variation in managerial characteristics and practices across managers. We next present preliminary evidence of how this variation coincides with line-level efficiency.

We select an example measure from 5 of the factors, omitting Personality and Demographics for the sake of brevity. We split the line-day productivity data at the median of each measure across line managers and construct box plots of efficiency for each subsample. Figures 4A through 4E, show that lines managed by supervisors exhibiting above median values of each measure of managerial quality achieve greater median productivity. For most of the measures the pattern is quite stark; while for some like digit span recall (a measure informing the Cognitive Skills factor) this simple comparison yields only a subtle pattern.

However, a comparison of the ranges conveys a more nuanced relationship. For some measures, such as internal locus of control (informing the Control factor) and autonomous problem-solving (informing the Autonomy factor), the minimum efficiencies are quite similar; while maximum efficiency and the interquartile range are notably higher for lines managed by supervisors with above median values of these measures. For other measures, such as tenure supervising current line (informing the Tenure factor) and active personnel management (informing the Attention factor), the most notable difference is that the minimum efficiency is much higher for lines managed by supervisors with above median values of these measures.

14. As discussed above, lines have between 1 and 3 permanent supervisors. While we have management characteristics for each manager, productivity data is common across managers of the same line. Co-supervisors generally share all production responsibilities, so it is only appropriate to match the productivity of a given line equally to each of the supervisors responsible. We will, however, account for this common mapping to productivity data in the bootstrap procedure by which we obtain errors for inference below.

TABLE 2
Managerial Quality Measures

	Mean	SD
Tenure		
Tenure Supervising Current Line	1.919	2.055
Tenure as Supervisor	4.779	3.117
Tenure in Garment Industry	10.074	4.411
Total Years Working	12.369	5.125
Demographics		
Egalitarianism	3.557	0.961
Demographic Similarity	4.872	2.340
Cognitive Skills		
Arithmetic (Number Correct)	11.517	3.706
Digit Span Recall	6.181	1.847
Arithmetic (% Correct of Attempted)	0.811	0.181
Control		
Internal Locus of Control	-5.000	4.334
Patience	2.107	1.269
Risk Aversion	3.148	1.472
Personality		
Perseverance	17.899	3.308
Conscientiousness	13.456	4.192
Self-Esteem	8.933	3.577
Psychological Distress	13.664	4.582
Autonomy		
Initiating Structure	47.644	5.712
Consideration	44.738	5.392
Self-Assessment	8.792	1.462
Autonomous Problem-Solving	-0.268	1.082
Identifying Production Problems	4.000	1.257
Attention		
Monitoring Frequency	4.846	0.415
Active Personnel Management	8.356	2.014
Issues Motivating Workers, Resistance	7.953	2.145
Efforts to Meet Targets	2.852	0.918
Lack of Communication	8.128	2.411
Pay		
Gross Salary (monthly)	14895.4	2024.6
Gross Pay with production bonus (monthly)	15079.7	2047.8

Note: Tenure variables are measured in years. Demographic similarity measures the similarities between the managers and the workers (range 0 to 6) and egalitarianism measures the preferences of the managers about the workers of the line (range 0 to 9). Digit span recall measures the number of correct digits a manager remember from a list of 12 numbers; arithmetic (number correct) counts the number of correct answers in a math test with 16 questions; arithmetic (% correct of attempted) is the ratio of the number of correct answers in a math test with 16 questions to the number of questions attempted. Locus of control is an index from -15 to 1; risk averse and patience are indices from 0 to 4. Perseverance is an index from 9 to 22; conscientiousness captures personality psychometrics from the Big 5 modules (range 3 to 20); self-esteem is an index from 1 to 16; psychological distress refers to Kessler's psychological distress scale (range 10 to 37). Initiating structure capture the degree to which a manager plays a more active role in directing group activities (range 30 to 50) and consideration capture a good rapport with subordinates (range 32 to 55); autonomous problem solving (range -3 to 2) and identifying production problem (range 1 to 7) measure the ability of the managers to identify and solve production problems alone; self-assessment measures one's evaluation of managerial quality relative to peers (range 5 to 10). Monitoring frequency is the number of rounds of the line to monitor production (range 2 to 5); efforts to meet targets is a composite index of dummy variables that measure the activities the supervisors reports engaging in to ensure that production targets are met (range 0 to 5); active personnel management is constructed analogously for activities related to reinforcing high level performance from star and under-performer workers (range 3 to 13); lack of communication measures the frequency of communication regarding production with both workers and upper level managers (range 3 to 18); issues motivating workers, resistance measures the frequency with which the supervisor reports issues motivating workers and overcoming resistance to initiatives and change (range 5 to 18).

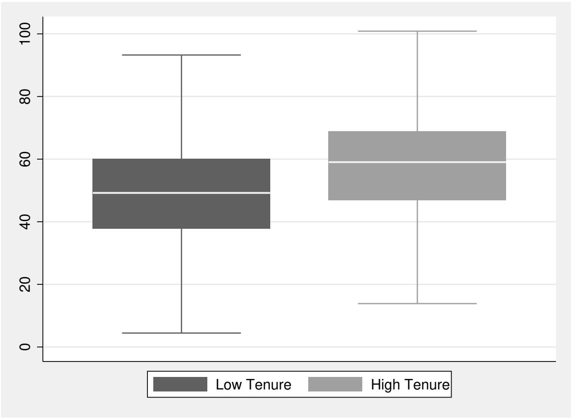

FIGURE 4A
Tenure Supervising Current Line

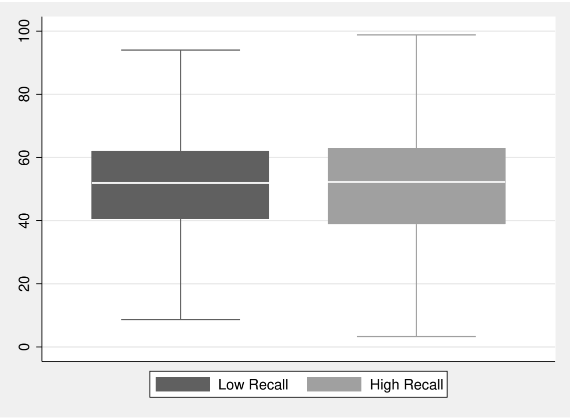

FIGURE 4B
Digit Span Recall

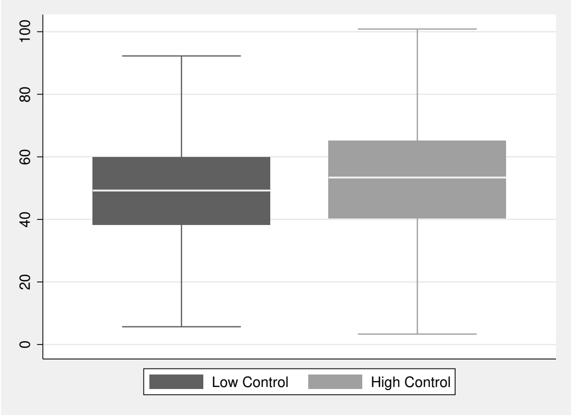

FIGURE 4C
Internal Locus of Control

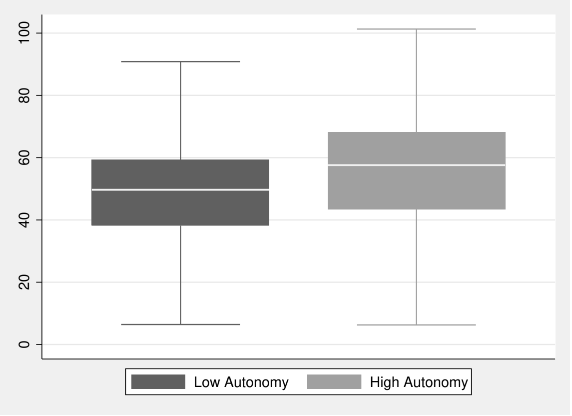

FIGURE 4D
Autonomous Problem-Solving

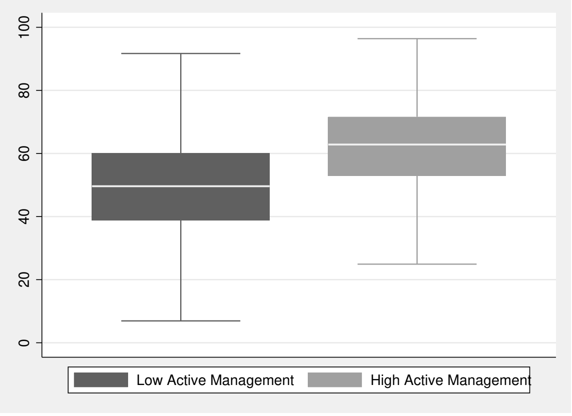

FIGURE 4E
Active Personnel Management

Note: In Figures 4A through 4E we split the line-day productivity data at the median of each measure across line managers and construct box plots of efficiency for each subsample (above and below median values).

In combination with the heterogeneity in average efficiency levels and dynamics across lines depicted in Figures 3A and 3B, these patterns suggest that variation in managerial characteristics and practices might explain a great deal of the variation in productivity both across lines and within lines over the course of the production run. In Figures A.1A through A.1E in the Appendix, we present learning curve analogues to the box plots presented in Figures 4A through 4E. That is, for each above and below median subsample of lines determined by each measure, we plot efficiency over the course of the production run. This evidence further indicates that line-level productivity indeed varies with managerial quality, and, furthermore, that some dimensions of managerial quality determine initial productivity more strongly while others determine the degree and rate of learning more strongly.

This preliminary evidence, of course, falls short of a formal investigation of these relationships. That is, ultimately we are interested in investigating the simultaneous, incremental contributions of each of these dimensions of managerial quality to variation in productivity. Such an exercise requires a more formal modeling of the learning function that allows for each quality dimension to flexibly contribute to both initial productivity and the rate of learning and acknowledges the noise and redundancy inherent in survey measures of managerial quality.

3. MODEL

3.1. *Learning Function*

In this section, we build a theoretical framework that formalizes the relationships implied by the preliminary results presented in the previous section. We start with a learning function with similar intuition and structure to that employed in Benkard (2000) and Levitt et al. (2013):

$$\log\left(y_{ijt}\right) = \alpha_i + \beta_i \log\left(E_{ijt}\right) + \psi_{J(i,t)} + \xi_t + \varepsilon_{ijt} \tag{3.1}$$

where $log\left(y_{ijt}\right)$ is the log daily efficiency of line $i \in \{1,...,N\}$, producing style $j \in \{1,...,J\}$ at period $t \in \{1,...,T\}$.[15] E_{ijt} is the experience that line i has in producing style j at date t in the current production run. α_i measures the initial level of productivity and β_i the rate of learning of line i. $\psi_{J(i,t)}$ is a fixed effect for the style the line was matched to at time t, and ξ_t is a time trend that is included in all specifications.[16] Finally, ε_{ijt}, is an idiosyncratic error term.[17]

We follow Benkard (2000) and assume that the experience, E_{ijt}, that line i has in producing style j at date t depends on the number of consecutive days spent producing that style in the current production run, $E_{J(i,t)}$, the cumulative experience with style j from previous productions run, P_{ijt}, and the number of days since line i produced style j before, D_{ijt}:

$$E_{ijt} = E_{J(i,t)} + \gamma P_{ijt}\left(1 - \delta D_{ijt}\right). \tag{3.2}$$

15. In additional results, omitted for the sake of brevity, we check that using log(quantity produced) on the left-hand side instead of log(efficiency) produces qualitatively identical patterns but with a smaller R-squared. Accordingly, we use log(efficiency) on the left-hand side as the preferred specification throughout. Given that efficiency is measured as the actual quantity produced exceeding minimum quality standards per worker-hour, it is also a closer analogue to the the defect rates and labor cost per unit used in previous studies (Levitt et al., 2013; Thompson, 2012).

16. The time trend is to account for any incidental serial correlation in productivity which may not reflect actual learning. In additional results, omitted for the sake of brevity, we check robustness to the inclusion of an additional control for days left to complete the order as a further check against this type confounding of incidental serial correlation with true learning, perhaps through "reference point" mechanisms.

17. Note that this function also matches closely to that used in and Benkard (2000) and Thompson (2001) with the factor allocations of capital ignored, given the fixed man-to-machine ratio in garment factories.

Here γ measures the contribution of previous stock learning (retention) and δ is the depreciation rate of previous stock learning (rate of forgetting) of line i. This functional form is meant to capture the degree to which learned productivity is retained from the past.[18]

Preliminary graphical evidence confirms a role for retention and forgetting in this empirical context. Figure 5A shows learning curves analogous to that depicted in Figure 2, but with the data split into first runs of a style on a line and subsequent runs. We see clearly that productivity gains accrued during first runs of a style are indeed retained, with lines starting at higher initial productivity levels and leaving less scope for additional learning. In Figure 5B we repeat the exercise, but with the sample of subsequent runs of the same style on a line further split by days elapsed since last run. We see that retained productivity gains from prior learning depreciates over the time elapsed before the line produces the same style again.[19]

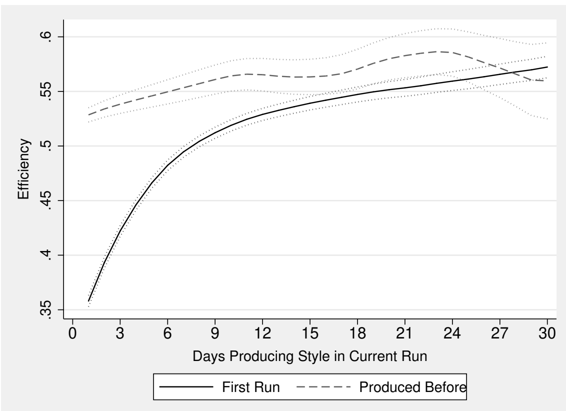

FIGURE 5A
Retention (Prior Days)

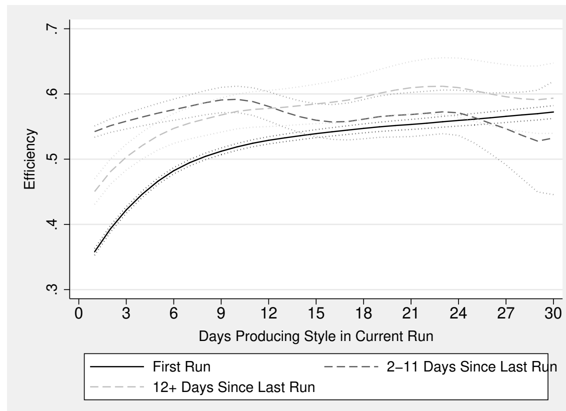

FIGURE 5B
Forgetting (Prior Days)

Note: Figures 5A and 5B depict the results of repeating the exercise from Figure 2, but further splitting previous runs by the number of days that have elapsed since the style was last produced. The fitted curves are the result of a lowess smoothed non-parametric estimation with a bandwidth of 0.25. Dotted lines represent 83% confidence intervals to emphasize significant differences between the two curves. Experience is trimmed at the 90th percentile in this graphical depiction to ignore outliers, but not from any regression analysis below.

3.1.1. Static Specification.

To investigate how important it is to consider learning dynamics when relating managerial quality to variation in line-day productivity, we also consider a model where the log daily efficiency depends only on a fixed effect representing the quality of the supervisor team of the production line, $\tilde{\alpha}_i$, in addition to the same fixed effect for the style the line was matched to at time t, $\tilde{\psi}_{J(i,t)}$, and time trend , $\tilde{\xi}_t$, as follows:

18. Note that γ and δ are assumed homogenous here across lines i. This is for parsimony, as estimating heterogeneity in these parameters would require restricting the sample to lines observed producing several styles multiple times. While most lines in the data indeed satisfy this requirement, losing even a few lines to this restriction would hamper the structural estimation of the contributions of each dimension of managerial quality. Preliminary investigation of heterogeneity along these dimensions revealed little informative value in relaxing this assumption. We discuss this in greater detail in section 4 below.

19. Figure A8 in the Appendix, presents a simple simulation to demonstrate that the functional form in equation (3.2) indeed yields the patterns depicted in Figures 5A and 5B.

$$\log(y_{ijt}) = \tilde{\alpha}_i + \tilde{\psi}_{J(i,t)} + \tilde{\xi}_t + \tilde{\varepsilon}_{ijt}. \tag{3.3}$$

Note that if the true data generating process (DGP) involves dynamics as reflected in equation (3.1), the estimation of a static model would yield distinct and incorrect parameters. For example, first days and weeks of orders are disproportionately represented in the data, as all orders have these early observations but many orders will not be long enough to have later days and weeks. Accordingly, initial productivity in an order will be over-emphasized in the average productivity of a line and in parameter estimates obtained from a static model. On the other hand, peak productivity achieved at the end of longer runs will be given appropriate attention when the learning dynamics are specified.

If these observations reflect large contributions from some dimensions of managerial quality, a static analysis will undervalue these dimensions. Counterfactual simulations and, as a result, generalizability will be impacted by misspecification of the DGP. Given that the patterns in the raw data shown in section 2 demonstrate an important role for dynamics, and even heterogeneity in these dynamics across lines, we prefer the dynamic model in equations (3.1) and (3.2). We carry the static model in (3.3) forward through the structural estimation for the sake of comparison and to demonstrate the value of the added complexity for interpretation and the counterfactual simulations below. Of course, we note that the value of the added complexity in the dynamic model is subject to the validity of the functional forms assumed.

3.2. *Parameterization of Relationship between Learning and Managerial Quality*

Note that the learning function in equation (3.1) differs primarily from those considered by the previous literature (Benkard, 2000; Levitt et al., 2013; Thompson, 2001) in that we allow for the parameters governing the shape of the learning curve (α_i, β_i) to vary across lines. This is done to reflect the graphical evidence presented in section 2.3 showing that the distributions of line-day productivity differ across lines supervised by managers with varying practices and characteristics in nuanced ways. We see that higher values of some dimensions of quality are associated with higher minimum efficiency, while others with higher maximum efficiency or both. Analogous figures in the Appendix confirm that these patterns are also reflected in learning curves, with some dimensions primarily yielding different initial productivities and others different rates of learning. However, we cannot tell from these simple exploratory graphs the functional form these relationships take.

Here we impose a structural form to arrive at an estimable relationship between managerial quality and each of the learning parameters. We follow the same approach for the alternative static model parameter ($\tilde{\alpha}$) as well. Specifically, we assume that there are k *latent* factors that describe managerial quality and that each of the learning parameters depends nonlinearly on these k factors, i.e.,

$$\exp(\iota_i) = f_\iota(\theta_{1,i}, \theta_{2,i}, \ldots, \theta_{k,i}) \tag{3.4}$$

where $\iota \in \{\tilde{\alpha}, \alpha, \beta\}$ for line $i \in \{1, \ldots, N\}$, and $\theta_{k,i}$ is the k-th quality factor. Note we assume that the functions for the supervisor/line effect from the static model ($f_{\tilde{\alpha}}$), initial level of productivity (f_α) and rate of learning (f_β) take the same set of underlying factors as arguments, but want to allow for the contributions of the factors to differ across these functions.

We assume that f_ι for $\iota \in \{\tilde{\alpha}, \alpha, \beta\}$ can be approximated by a Constant Elasticity of Substitution (CES) function. The CES form allows us to explore the degree of complementarity or substitutability between the factors included in the function for each learning parameter. That is, we assume that f_ι takes the following functional form,

$$\exp(\iota_i) = A_\iota[\lambda_{\iota,1}\theta_{1,i}^{\rho_\iota} + \lambda_{\iota,2}\theta_{2,i}^{\rho_\iota} + \cdots + \lambda_{\iota,k}\theta_{k,i}^{\rho_\iota}]^{\frac{1}{\rho_\iota}} \tag{3.5}$$

where $\lambda_{\iota,k} \geq 0$ and $\sum_k \lambda_{\iota,k} = 1$ for $\iota \in \{\tilde{\alpha}, \alpha, \beta\}$ and line $i \in \{1, ..., N\}$. Note that any of the factors can be irrelevant in any of these functions when $\lambda_{\iota,k} = 0$. ρ_ι determines the elasticity of substitution between the latent factors, which is defined by $\frac{1}{1-\rho_\iota}$, and A_ι is a factor-neutral productivity parameter. Under this technology, $\rho_\iota \in [-\infty, 1]$; as ρ_ι approaches 1, the latent factors become perfect substitutes, and as ρ_ι approaches $-\infty$, the factors become perfect complements. Note we assume a common functional form across the learning parameters $\iota \in \{\tilde{\alpha}, \alpha, \beta\}$, but we allow the loadings for each latent factor k ($\lambda_{\iota,k}$) and the degree of complementarity (ρ_ι) to differ across learning parameters and the management parameter from the alternative static model.

We acknowledge that the CES form imposed is a compromise, allowing for nonlinearities between different dimensions of quality, but in an admittedly restricted way. That is, one might expect that complementarities or substitutabilities between the *k* dimensions of quality are unlikely to be strictly symmetric or global in the way imposed by the assumed CES form. The quantity of dimensions studied (*k*) makes estimating heterogeneity in the sign or magnitude of pair-wise interactions empirically challenging, despite the relatively large number of lines and managers in our data.[20]

We explored the possibility of a nested form emphasizing key pair-wise interactions which the data might suggest would differ from others or dominate, but found no clear evidence to motivate a more complicated functional form. As such, we adopt the global non-linearity form of the standard CES as a compromise and interpret the complementarity/substitutability parameter as indicative of interactions between dimensions of managerial quality and valuable in allowing for flexibility in counterfactual simulations below. We leave the study of specific interactions between dimensions of managerial quality for future work, likely in an empirical context in which the number of production teams and supervisors is larger relative to the number of dimensions of managerial quality being studied.

4. EMPIRICAL STRATEGY

Having adapted the canonical learning function to allow different dimensions of managerial quality to flexibly determine the shape of the learning curve, we next develop our strategy for estimating these relationships in the presence of measurement error. Remember that our goal is to be able to estimate equation (3.5) for $\iota \in \{\tilde{\alpha}, \alpha, \beta\}$. However, to do so, we must first recover $\tilde{\alpha}_i$, α_i, and β_i, for the LHS of equation (3.5), and also extract the k latent factors $\theta_{k,i}$ for the supervisors of each line i from the management survey data.

Accordingly, our empirical strategy consists of three steps. First, we estimate equation (3.1) and (3.3) to recover $\{\alpha_i, \beta_i\}$ and $\tilde{\alpha}_i$, respectively, for each line $i \in \{1, ..., N\}$ using ordinary least squares. Second, we follow Cunha et al. (2010) Attanasio et al. (2015), and Attanasio et al. (2015) in estimating a nonlinear latent factor measurement system using the data from our managerial survey. This step allows us to recover information about the joint distribution (approximated as a mixture of two normals) of k latent factors (θ_k) underlying the multitude of noisy survey measures and the learning parameters estimated in the first stage $\{\alpha_i, \beta_i\}$ using maximum likelihood and minimum distance. We finally draw a synthetic dataset from this joint distribution and estimate equation (3.5) for $\iota \in \{\alpha, \beta\}$ using nonlinear least squares and bootstrap the entire procedure as

20. Note pair-wise interactions between the 7 latent factors yield 21 interactions, meaning 28 parameters including the 7 main effects. Needing to estimate this for both α and β means we would be trying to estimate 56 parameters from 120 production lines.

described below to obtain the error distribution. We also perform the entire procedure using the static manager fixed effect parameter estimated in the first step, $\tilde{\alpha}_i$.

4.1. *First Stage: Productivity Dynamics*

4.1.1. Homogeneous Learning Function. We start by estimating the conventional model of learning-by-doing represented by equation (3.1) but with homogeneous parameters for α and β. At first, we restrict experience to only the number of consecutive days spent producing a style in the current production line (i.e., ignoring the stock evolution described in equation (3.2)). To account for worker composition, we first regress the log of efficiency and the log of the number of consecutive days in the current production line on worker fixed effects. We get the residuals from each regression, and estimate equation (3.1) by ordinary least squares using different sets of cross-sectional and temporal fixed effects.[21] In particular, we include style fixed effects to account for variation in productivity due to complexity of the style and size of the order, as well as year, month and day of the week fixed effects, to account for common seasonality and growth in productivity across lines.

Next, we estimate learning-by-doing with homogeneous learning parameter, assuming that stock experience evolves as described in equation (3.2). That is, the current experience depends on the number of consecutive days spent producing a style in the current production run, the cumulative experience retained from previous production runs, and the depreciation of this stock of experience over the days since the line last produced the same style. This model matches the specification used in previous studies of learning-by-doing (Benkard, 2000; Levitt et al., 2013; Thompson, 2001) and is represented by equations (3.1) and (3.2) with homogeneous parameters for α and β. We perform this estimation by nonlinear least squares.[22] These estimations serve to validate that the patterns observed in Figures 2, 5A, and 5B indeed persist in a more formal regression framework and that the functional form in equations (3.1) and (3.2) fit the patterns in the data well. We also use these estimations to demonstrate that the patterns are robust to controls for worker composition and qualitatively similar when allowing for the stock of experience to reflect previous production runs of the same style.

4.1.2. Heterogeneous Learning Functions. Next, we estimate the learning function from equation (3.1) as it is written, allowing for initial levels of productivity and rate of learning to vary across lines. Using the homogenous parameters estimated from equation (3.2), we construct a measure of current experience, $\hat{E}_{ijt}$, that accounts for number of consecutive days spent producing a style in the current production run, the cumulative experience retained from previous production runs, and the depreciation of this retained cumulative experience over the days since the line last produced the same style. We restrict γ and δ to be homogenous for the sake of parsimony, as estimating heterogeneity in these parameters would require restricting the sample to lines observed producing several styles multiple times. While most lines in the data indeed satisfy this requirement, losing even a few lines to this restriction would hamper the structural estimation of the contributions of each dimension of managerial quality described below.[23] As

21. We follow the approach used by Adhvaryu et al. (2021) in a nearly identical empirical setting. They note that including the full set of worker fixed effects is sufficient to recover line or supervisor contributions to productivity.

22. We again account for worker composition by first regressing log of efficiency on worker fixed effects and using the residual in the estimation of the learning function.

23. Relaxing this assumption for γ and δ in the first stage explains only an additional 6-7% of the variation in productivity, which is only roughly a tenth of the contribution of the heterogeneity in α and roughly a third of that of β.

in the homogenous estimation above, to account for worker composition, we first regress log of efficiency and log of $\hat{E}_{ijt}$ on worker fixed effects. We use the residuals from each regression to estimate equation (3.1) and recover α_i and β_i for each line $i \in \{1,...,N\}$. Finally, we do the same to estimate equation (3.3) and recover $\tilde{\alpha}_i$, for each line $i \in \{1,...,N\}$.

4.1.3. Identification of First Stage Parameters. The style fixed effect in addition to the line-specific learning parameters being estimated amounts to a two-way fixed effect model of lines matched to styles. This two-way fixed effect model is analogous to the worker-firm sorting model studied by Abowd et al. (1999) (also known as AKM).[24] Accordingly, we must address, as they do, the potential obstacles to identification of the parameters of interest due to any possible endogenous sorting in the match between lines and styles in the data and/or any non-separability between these to effects in the determination of line-day-level log efficiency.

First, note that to be able to the identify the line and style fixed effects separately, lines must be observed producing different styles for multiple production runs during the sample period, and each style should be observed being produced by multiple lines (not necessarily contemporaneously). Second, identification is possible only within a group of lines and styles that are connected. A group of lines and styles are connected when the group comprises all the styles that have ever matched with any of the lines in the group, and all of the lines at which any of the styles have been matched during the sample period. Third, we must assume (and check to the best of our ability) that the probability of a style being produced by a certain line is conditionally mean independent of contemporaneous, past, or future shocks to the line. Fourth, we must assume that there is no complementarity or non-separability between lines and styles in the DGP for productivity.

The third and fourth assumptions are quite strong. For example, if the firm is aware of the heterogeneous productivity dynamics described in section 2.3, it stands to reason that the firm would consider these differences in productivity levels and dynamics when allocating styles so as to optimize overall productivity. However, if either the firm does not actively measure and analyze these differences in dynamics or the underlying managerial characteristics, or the firm is incapable of practicing this type of optimal allocation of styles to lines due to difficulty in forecasting the arrival of future orders and/or a high cost of leaving lines vacant to await optimally matched orders in the future, then we might expect that assumptions 3 and 4 might actually hold in the data. In the next section, we present checks of these assumptions in our context.

4.1.4. Test for Endogenous Assignment of Styles. To establish the validity of identification assumptions, we need to check for endogenous sorting of styles/orders to managers/lines and for non-separability between line and style contributions in the determination of productivity.[25] We might imagine that some coarse insights might be gleaned from less rigorous

24. We have a two-way FE model in which the lines and styles map to the firms and workers, respectively, in the context of the AKM model.

25. We might also be concerned, as discussed in section 2 above, about the possibility of sorting of workers to lines. Adhvaryu et al. (2021) establish that including the full set of worker fixed effects in log efficiency specifications is sufficient to recover line/manager contributions to productivity net of the effects of worker composition of each line on each day. Accordingly, we follow their approach by projecting off worker fixed effects from line-day measures of efficiency and experience prior to estimating the first stage. We also performed worker composition balance checks in which we compare different characteristics of the workers (efficiency, skill grade, salary, age, tenure, gender, language, and migrant status) across high and low-type managers defined by the 26 different measures included in the measurement system. Only 29 out of 234 differences are statically significant with significant differences spread across various manager characteristics. Tests of joint significance cannot reject balance overall. We omit these tables for the sake of brevity.

measurement and analysis which might allow the firm to optimize the allocation of styles to lines. Such dynamic optimal assignment would, however, require both predictability of future orders and a willingness to delay the start of an order and leave some lines vacant for some periods of time to achieve a more optimal match of style to line. Ordering and production scheduling in the export garment industry is by all accounts "just in time," given the importance of seasonal trends and the high cost of retail inventory storage. We find no evidence that lines are ever left vacant.[26]

This evidence is all consistent with a limited predictability of future orders and a high cost of slackness as communicated by factory management. However, to check empirically that indeed there is quasi-random style assignment to production lines, we begin by checking that the length of the order (in number of days) are balanced across the managerial characteristics used in our latent factor measurement system. The comparisons presented in Table A2 and show very few (2 of the 26) significant differences. Moreover, joint tests fail to reject balance overall.[27]

We also follow Card et al. (2013) and perform a series of tests for endogenous mobility of production lines across styles. We begin by conducting an event study around moves of lines across styles of differing average productivity to assess the extent to which moves might be systematically driven by productivity shocks or by sorting on a line-style match-specific component of log efficiency. We focus on the styles that are produced in more than one production line, and then rank them in terms of quartiles of average efficiency achieved in producing that style.

Figure A2 plots on the y-axis the average weekly log-efficiency of production lines across two consecutive orders, by quartiles of the average efficiency of the styles on which the line is working.[28] If style assignments are not driven by match-specific components, lines switching to styles with higher average efficiency, will achieve higher efficiency on average, and lines switching to styles with lower average efficiency will achieve lower efficiency, after the move. Moreover, as discussed in Card et al. (2013), if the moves are conditionally mean independent of the match-specific component, then the gains from moving from style X to style Y should be equal in magnitude to the losses from moving from style Y to style X. That is, gains and losses for movers should be symmetric.

We see in Figure A2 clear evidence of precisely this pattern. We also see that lines switching from a style in the highest quartile to another style in the highest quartile experience close to zero change in productivity, and the same is true for lines switching between two styles in the lowest quartile. These results are consistent with the absence of an average "premium" for styles produced in more than one line. We present a full symmetry test across all potential combinations of origin and destination styles (ranked by quartiles of average efficiency) in Figure A3. The patterns in Figure A3 further confirm that the gains from moves up and the losses from moves down are remarkably symmetric, consistent with the assumption of conditional mean

26. One other concern might be that when a given line is underperforming on a current style, upper management might reassign the style to another higher productivity line. We were told anecdotally that this rarely happens if at all largely because the steep learning curve we document here means that even a relatively more productive line might be quite unproductive on the newly reassigned style even compared to the relatively unproductive line that is underperforming on it currently. Indeed, we find in the data that on less than 2% of days in which a style changes on one line do we observe another line producing the same style, which is itself an upper bound statistic for the degree of this "reassignment" as most of these days will reflect simply distinct orders of the same style being produced on different lines according to plan rather than true "reassignment" due to underperformance.

27. The incidental individual differences do not appear to systematically match to the pattern of findings across dimensions of managerial quality presented and discussed below.

28. This is computed 8 to 12 days ($Period = -2$) and 3 to 7 days ($Period = -1$) before the end of the first order, and 3 to 7 days ($Period = 1$) and 8 to 12 days ($Period = 2$) of the second consecutive order, as reported on the x-axis. To improve readability, we only report production lines switching from the top style-quartile in terms of average efficiency (quartile 4) or the bottom style-quartile of average efficiency (quartile 1).

independence of style-line assignments and an absence of line-style match effects in the DGP for line-day-level efficiency.

Additionally, to further check if line-style match effects are important determinants of productivity, we compare the adjusted R^2 from the estimation of equation (3.3) with the adjusted R^2 from a fully saturated model with dummies for each line-style combination. Table A3 in the Appendix shows that the fit improvement from the fully saturated model is minimal, suggesting that match-specific components play a minor role in determining productivity. Finally, in Figure A4 in the Appendix, we plot the average residuals from the estimation of equation (3.3), by deciles of the estimated line and style fixed effects. The average residuals are very small for all groups, and again we see no systematic pattern, consistent with match effects not being quantitatively important.

4.1.5. Addressing Limited Mobility Bias. Remember that identification of the line and style fixed effects requires observing lines producing different styles, and observing styles being produced by multiple lines. If the number of style "movers" across lines is limited, we cannot separately identify the style and line fixed effects. This issue is referred to as "limited mobility bias" in the literature (Abowd et al., 2002; Andrews et al., 2008, 2012). Table 1 shows, in our data, 41% of styles are produced at more than one line, and the median of the number of different styles produced per line is 27. As a comparison, the share of worker movers across firms is around 12% in Andrews et al. (2012), 25% in Card et al. (2016), and around 35% in Alvarez et al. (2018). Accordingly, concerns regarding limited mobility bias are limited in our setting. Nevertheless, as a robustness check, we perform the covariance shrinkage method proposed by Best et al. (2019), who extend shrinkage methods (e.g., Kane and Staiger (2008), Chetty et al. (2014)) to explicitly account for the correlation between the estimation error of the two vectors of fixed effects.[29] These results are reported in Tables A9 and A10 in the Appendix and are quite similar to our main results.

4.1.6. Monte Carlo Simulations. Despite reassuring evidence from the above now standard checks in the literature of the identifying assumptions underlying the first stage two-way fixed effects model, we also assess if there is any composite bias in our estimation due to limited or endogenous mobility of lines across styles using the Monte Carlo experiment proposed by Abowd et al. (2004) which relies on the observed match pattern between lines and styles. We first estimate the model in equation (3.1) and preserve all the observed characteristics, line and style *identifiers*, and the realized line-style match of each observation. We draw for each style a *style effect*, and for each line an initial productivity and rate of learning (i.e., our proposed decomposition of the *line effect*) from normal distributions with the same moments as estimated for the distributions of the line and style parameters in the first step.[30] Finally, we construct simulated productivities from the

29. We use bootstrap estimation to construct the variance of our manager and style fixed effects. To account for the covariance of the estimation errors of the line and style fixed effects, we follow the shrinkage approach proposed by Best et al. (2019): let $\hat{\Theta}$ be a vector with the estimated style and line fixed effects. Then, the shrinkage matrix $\Lambda*$ is defined as $\operatorname{argmin}_{\Lambda} \mathbf{E}\left[\left(\Theta - \Lambda\hat{\Theta}\right)\left(\Theta - \Lambda\hat{\Theta}\right)\right]$. That is, we find the weights Λ that minimize the expected mean squared error of the prediction of the linear combination of worker and manager fixed effects. We then "shrink" the estimated style and line fixed effects by multiplying them by such weights.

30. That is, we compute the mean and standard deviation of the estimated line effect parameters (i.e., initial productivity and rate of learning) and style effects. We simulate the new lines and styles effects using these moments. Note that the line *effect* (i.e., initial productivity and rate of learning) and the style *effect* for each line is drawn independently of the other, such that the match-specific effects are forced to be null.

randomly drawn line and style parameters and a randomly drawn idiosyncratic error term.[31] We then estimate the model again using this simulated data. We repeat the procedure 10,000 times, and compute the percentage mean bias in absolute value between the original estimates from the real data and these bootstrapped estimates from the simulated data for the coefficients of interest (α_i and β_i). This procedure measures the degree to which limited mobility and/or endogenous matching of lines to styles are biasing estimates of equation (3.1). Table A5 in the Appendix shows minimal bias (well less than 1%), indicating that the sorting assumptions imposed in the first stage estimation are valid in the data.

4.2. *Second Stage: Latent Factors of Managerial Quality*

We do not directly observe $\theta_{k,i}$. Instead, we observe a set of measurements that can be thought of as imperfect proxies of each factor with an error. We adapt from Cunha et al. (2010) a non-linear latent factor framework that explicitly recognizes the difference between the available measurements and the theoretical concept used in the production function. We set the number of the latent factors to $k=7$, comprised of the following: Tenure, Demographics, Cognitive Skills, Control, Personality, Autonomy, and Attention. As discussed in section 2.3, we use the original survey module delineations and exploratory factor analyses, following Attanasio et al. (2015,?) and Cunha et al. (2010), to map the full set of survey measures to these seven factors, each corresponding to dimensions of managerial quality previously proposed and studied in the literature. That is, we let both the intuition of the modules and the data itself determine which are the distinct factors and which measures map to each factor.

Let $m_{l,k}$ denote the lth available measurement relating to latent factor k. Following Cunha et al. (2010) and Attanasio et al. (2015), we assume a semi-log relationship between measurements and factors such that

$$m_{i,l,k} = a_{l,k} + \gamma_{l,k} \ln\theta_{i,k} + \varepsilon_{i,l,k} \tag{4.6}$$

where $\gamma_{l,k}$ is the factor loading, $a_{l,k}$ is the intercept and $\varepsilon_{i,l,k}$ is a measurement error for factor $k \in K \equiv \{T, D, Ctrl, Cog, P, R, Aut, Att\}$ (Tenure, Demographics, Cognitive Skills, Control, Personality, Autonomy, and Attention) and measure $l \in \{1, 2, ..., M_k\}$. Thus, for each k we construct a set of M_k measures.

For identification purposes, we normalize the factor loading of the the first measure to be equal to 1 (i.e., $\gamma_{1,k} = 1$ for $k \in K$). Similarly, log-factors are normalized to have mean zero, so a_{lk} is equal to the mean of the measurement. Finally, $\varepsilon_{i,l,k}$ are zero mean measurement errors, which capture the fact that the $m_{i,l,k}$ are imperfect proxies. Three assumptions regarding the measurements and factors are required for identification. First, we assume that the latent factor and the respective measurement error are independent. Second, we assume that measurement errors are independent of each other. Finally, we assume that each measure is affected by only one factor.[32]

31. We first assume that the errors are i.i.d. across lines and time, and then relax this assumption by using the autocorrelation structure estimated of the residuals from the estimation on the real data in the first step.

32. This assumption can be relaxed to allow some subset of measures to inform more than one factor; however, in our setting, these cross-factor loadings are not well-motivated, as factors come from distinct modules of the survey which were designed to capture different aspects of managerial quality. For identification of the system, we need at least two dedicated measures per factor and at least one measure for each factor conditionally independent of the other measures. See Cunha et al. (2010) and Attanasio et al. (2015). Note as discussed in 2.3 that in exploratory analyses across pooled sets of measures across modules we find some correlations; however, we always assign the measure to the factor for which its loading is strongest. Note that the factors obtained can be correlated with each other and indeed do appear to be in the final results as shown in the Appendix. Accordingly, this assumption preserves the interpretation of each factor while not restricting that measures assigned to different factors be unrelated.

Note that the estimation of (3.5) requires the construction of a synthetic dataset from the joint distribution of management factors and estimated learning parameters. We follow Attanasio et al. (2015) and augment the set of latent factors with $\hat{\alpha_i}$ and $\hat{\beta}_i$, estimated in the first stage, and the average of the log of supervisor pay, w_i, for each line i.[33] As we explain later in Section 6, we are able to recover α_i and β_i for 120 lines, which is the largest connected set.[34]

Finally, we assume that the learning parameters from the first stage and the log of supervisor pay are measured with no error.[35] Let $\boldsymbol{\theta}_i \equiv \left(\theta_{i,1},\theta_{i,2},\theta_{i,3},\theta_{i,4},\theta_{i,5},\theta_{i,6},\theta_{i,7},\exp(\hat{\alpha_i}),\exp\left(\hat{\beta_i}\right),\exp(w_i)\right)$ and $\boldsymbol{\theta} \equiv (\boldsymbol{\theta}'_1,...,\boldsymbol{\theta}'_N)$, thus we can express the *extended* demeaned measurement system in vector notation as,

$$\tilde{M} = M - A = \Lambda \ln(\boldsymbol{\theta}) + \Sigma_\varepsilon \varepsilon \tag{4.7}$$

where M is $\left(\sum_{k\in K} M_k \times N\right)$ matrix of measures, Λ is $\left(\sum_{k\in K} M_k \times K\right)$ matrix of factor loadings, ε is a $\left(\sum_{k\in K} M_k \times N\right)$ matrix of measurement errors, and Σ_ε is a $\left(\sum_{k\in K} M_k \times \sum_{k\in K} M_k\right)$ diagonal matrix with the standard deviation of the measurement error defined before.[36]

In order to capture non-separabilities between the factors in determining the learning parameters as well as to allow for correlations between the factors, we follow Cunha et al. (2010) and Attanasio et al. (2015) in assuming that the joint distribution of the log latent factors, $f(\cdot)$, follows a mixture of two normals,

$$f(\ln\boldsymbol{\theta}) = \tau f^A(\ln\boldsymbol{\theta}) + (1-\tau) f^B(\ln\boldsymbol{\theta}) \tag{4.8}$$

where $f^n(\cdot)$ is the joint CDF of a normal distribution with mean vector, μ_n, and variance covariance matrix, Σ^n, and mixture weight, $\tau \in [0,1]$, for $n \in \{A,B\}$.[37] Finally, we assume that the log-factors have mean zero, i.e.,

$$\tau\mu^A + (1-\tau)\mu^B = 0 \tag{4.9}$$

Note that if ε is normally distributed, the distribution of the observed measurements is

$$\mathcal{F}(m) = \tau \cdot \Phi(\mu_{m_A}, \Sigma_{m_A}) + (1-\tau) \cdot \Phi(\mu_{m_B}, \Sigma_{m_B}) \tag{4.10}$$

where,

33. We use total compensation of the supervisor for the month which includes the monthly salary for November 2014, the month in which the management survey was completed, and any production bonus associated with the productivity of the line.

34. This largest connected set contains 98.61% of the raw data available.

35. This assumption with respect to the pay measure is similar to that imposed by Attanasio et al. (2015) in their extended measurement system. With respect to the learning parameters and manager effect from the static model, we are including constructed variables in our second stage. From the validity of the identification in the first stage, we regard the error remaining in the constructed variables ($\hat{\alpha_i}$ and $\hat{\beta}_i$) to be near 0 as $T \to \infty$. Relaxing this assumption would require multiple measures for each of the learning parameters which we do not have. Nevertheless, we bootstrap all three stages of the procedure to construct errors for inference as described below and therefore do not rely heavily on this assumption.

36. As mentioned before, we assume that learning parameters and the log of pay are measured with no error. This implies that the corresponding factor loadings are set equal to one in Λ, and the corresponding standard deviations of the error in Σ equal to zero.

37. The departure from the joint normality assumption underlying traditional factor analyses is important, otherwise the log of the production function would be linear and additively separable in logs (i.e., Cobb-Douglas, as discussed in Attanasio et al. (2015)) and the factors would be assumed orthogonal. The mixture of normals allows for a relaxation of this assumption in a tractable and quite flexible way (Attanasio et al., 2015; Cunha et al., 2010).

$$\mu_{m_A} = \Lambda \mu_A \tag{4.11}$$

$$\mu_{m_B} = \Lambda \mu_B \tag{4.12}$$

$$\Sigma_{m_A} = \Lambda' \Sigma_A \Lambda + \Sigma_\varepsilon \tag{4.13}$$

$$\Sigma_{m_B} = \Lambda' \Sigma_B \Lambda + \Sigma_\varepsilon \tag{4.14}$$

Estimation in this second stage proceeds in three steps. First, we construct the set of measures for each latent factor by matching the appropriate survey modules to each of the seven dimensions of quality previously studied in the literature, as discussed in section 2.3. Second, we use maximum likelihood to estimate an unconstrained mixture of normals for the distribution of measurements.[38] Using equations (4.9) through (4.14) as restrictions, we perform minimum distance estimation to recover $\mu^A, \Sigma^A, \mu^B, \Sigma^B$. Finally, we draw a synthetic dataset from the joint distribution of the learning parameters (as well as the log of pay) and factors of managerial quality to produce data for both the LHS and RHS of equation (3.5). We repeat the steps described in the previous paragraph, using the same set of measures for each latent factor, augmenting the set of latent factors with the estimated supervisor/line effect from the static model, $\hat{\tilde{\alpha}}_i$, and the average log of supervisor pay, w_i, for each line i.[39]

4.3. *Third Stage: Contributions of Managerial Quality to Productivity Dynamics and Pay*

Remember that our goal is to estimate equation (3.5) for $\iota \in \{\alpha, \beta\}$ (and alternately $\iota \in \{\tilde{\alpha}\}$). We first recover the learning parameters (initial level of productivity and rate of learning) for the LHS of equation (3.5) for each line by estimating the line-specific learning function in equation (3.1) using ordinary least squares. Second, we estimate a latent factor measurement system similar to Cunha et al. (2010) and Attanasio et al. (2015) and recover the joint distribution of the latent factors and the learning parameters obtained in the first stage. That is, from the full set of error-ridden survey measures we observe, we recover the RHS of (3.5). This procedure allows us to construct a synthetic dataset of the factors (RHS) and the learning parameters (LHS). Finally, in the third stage, we estimate jointly the system of equations (3.5) for $\iota \in \{\alpha, \beta\}$ using nonlinear least squares. We also repeat this last step with log of mean supervisor pay on the LHS instead, keeping the functional form and set of factors taken as arguments on the RHS the same.

Similarly, we recover the supervisor/line effect by estimating static version of the two-way fixed effect model in equation (3.3) using ordinary least squares. We estimate a latent factor measurment system and recover the joint distribution of the latent factors and the static supervisor/line effect. We construct a synthetic dataset of the factors (RHS) and line fixed effects (LHS), and estimate equations (3.5) for $\iota \in \{\tilde{\alpha}\}$ using nonlinear least squares.

We bootstrap the entire procedure — all three stages — from the top. To construct the bootstraps standard errors, we follow Best et al. (2019) by constructing residuals for the first stage and randomly resampling the residuals, stratifying by line-style pair to preserve the match structure of the observations. We then re-estimate the line and style fixed effects, and estimate

38. We use EM algorithm and k-means clustering to select the initial values with uniform initial proportions. We replicate the procedure 10,000 times and select the model with largest loglikelihood.

39. Again, we assume that the estimated average productivity from the first stage and the log of supervisor pay are measured without error.

the second and third stage. We repeat this procedure 200 times, and use the means and standard deviations of estimates across replications for inference.[40]

Note that identification of the LHS of equation (3.5) is achieved by the AKM assumptions stated, and to the degree possible tested, in the discussion of the first stage above. However, interpreting the estimated parameters on the RHS of equation (3.5) as causal contributions would require stronger assumptions regarding exogenous allocations of the dimensions of skill ($\theta_{k,i}$) across supervisors of different production lines. Such an assumption is unlikely to be true and as such we treat this third stage as a decomposition of the causal line-level contributions to productivity (α,β) into factor-specific components. That is, we take a structural approach to inference by explicitly modeling and recovering the correlation between the assignment of factors across supervisors (presented in Table A4 in the Appendix). We then use this recovered correlation structure and the imposed functional forms (i.e., CES) in counterfactual simulations.[41]

5. RESULTS

In this section, we present and interpret the results of the estimation described in section 4. We first report estimates from the learning function given by equations (3.1) and (3.2) assuming homogeneous parameters across lines to verify that the pattern observed in Figure 2 is well fit and robust to accounting for worker composition and prior experience with the same styles. We then estimate the learning function with heterogeneous parameters, recovering α_i and β_i for each production line. Next, we discuss the estimated signal in each of the measures used in the latent factor measurement system to recover the underlying dimensions of managerial quality. Then, we report estimates from equation (3.5) for $\iota \in \{\alpha, \beta\}$, as well as $\iota \in \{\tilde{\alpha}\}$ as indicated by the alternative static model and interpret differences in the results of these two models.

To aid in the interpretation of these estimates, we perform simulations to investigate how productivity changes with increases in each of the dimensions of managerial quality, leveraging both the estimated correlations between the factors and their functional combination in determining productivity dynamics. We weight initial productivity and rate of learning according to how much each contributes to average productivity in the sample and first pin the experience profile of the simulated line to the observed mean. Then, we repeat the exercise assuming the line sees only short one-off orders or long repeat orders, alternately, in order to highlight which dimensions of quality should be emphasized in each case. Finally, we use our procedure to investigate the relationship between the latent factors for managerial quality and the observed pay of supervisors, and perform analogous simulations to recover pass through of productivity contributions of each dimension of managerial quality to pay.

5.1. *First Stage: learning parameters*

Table 3 presents the results of the learning function in equation (3.1) with homogeneous learning parameters. In the left two columns, we ignore prior experience with the same style, estimating the log linear model by ordinary least squares; while in the right two columns we allow for the stock of experience to evolve as given in equation (3.2), estimating the resulting model by nonlinear least squares. While all specifications include style fixed effects, results in columns 2 and 4 also account for worker composition by projecting off the full set of worker fixed effects from the

40. Note that when performing this from the first stage results, which are obtained from line-day productivity records, multiple supervisors assigned to the same line will be treated as a cluster or block.

41. As is always the case in these structural approaches, the interpretation of this inference is subject to these imposed modeling and functional form assumptions.

outcome.[42] Table 3 shows that the estimated learning rate is between 0.105 and 0.135, implying that productivity will increase 30 to 40% over the course of about 3 weeks of producing the same style, consistent with Figure 2. Furthermore, the results indicate that accounting for worker composition and prior experience with the same style do not substantially alter this empirical relationship. Nevertheless, we use the most rigorous version of the model represented in the right most column throughout the remaining analysis below.

TABLE 3
Learning (Experience in Days)

	Log(Efficiency) (Actual Production/Target Production)			
	Ignoring Previous Experience		*Full Model*	
β	0.119***	0.132***	0.135***	0.104***
	(0.008)	(0.009)	(0.007)	(0.006)
γ			0.516***	0.315***
			(0.066)	(0.052)
δ			0.004***	0.004***
			(0.00003)	(0.00007)
Observations	49,976	49,976	49,976	49,976
Additional Controls	Style FEs	Style and Worker FEs	Style FEs	Style and Worker FEs

Note: Standard errors are clustered at the line level.

Next, we estimate equation (3.1) for each line by ordinary least squares to recover α_i and β_i. We use the homogenous γ and δ from the right most column of Table 3 to construct the stock of experience for each line-style-day observation from a combination of experience on the current run, previous cumulative experience with the same style, and days since the same style was last produced on the line.[43] We also, alternately, estimate the static model in equation (3.3) for each line to recover $\tilde{\alpha_i}$.

42. In estimating these robustness specifications, we make two implicit assumptions about the data generating process: 1) that the worker fixed effects enter the learning function linearly (in logs) and are additively separable from the non-linear portion of the function (in column 4), and 2) that the worker fixed effects are conditionally mean independent of the other terms in equation 1. Under these assumptions, it is appropriate to effectively partition the estimation (although ours is not a strict partition regression, as we collapse the data to the line-day level after projecting off the worker fixed effects but before estimating equation 1). Given that the estimates in Table 3 do not vary substantially across specifications (without projecting off worker FE from the outcome in columns 1 and 3 as compared to projecting off worker FE from the outcome in columns 2 and 4), we do not explore relaxing this assumption.

43. As we discuss above, we use the largest connected set representing 98.61% of the available data. If we were to allow γ and δ to also vary across lines, we would have to further restrict the sample to lines observed producing multiple styles multiple times. Though the majority of lines satisfy this requirement, losing even a few lines would impact the degrees of freedom in the second and third stages below. Furthermore, variance decomposition revealed that heterogeneity in γ and δ incrementally explained only 6-7% of the variation in line-day level productivity. Finally, a comparison of estimates of β across columns in Table 3 indicates that even ignoring prior experience altogether would not substantively impact the study of rates of learning.

In Figures 6A and 6B we plot the rank of each line in terms of $\hat{\alpha}_i$ and $\hat{\beta}_i$, respectively, against the rank in terms of $\hat{\hat{\alpha}}_i$. We see in Figure 6A that the static model yields a supervisor/line contribution to productivity that aligns well though not perfectly with the initial productivity parameter from the model with learning dynamics. However, Figure 6B shows that the static model parameter does not appear to capture the heterogeneity in rates of learning across lines.[44] This pattern is in line with the intuition we describe in section 3.1.1 the static model will overweight early days in each order as compared to later peak productivity observations.

How this impacts the study of the relationship between managerial quality and productivity will depend on how much this quality works through the rate of learning which remains to be seen. In Table A1 in the Appendix we report a comparison of the decomposition of the variation in productivity between the learning model and the static model. We see that the static model overstates the explanatory power of style fixed effects and time controls in addition to the line intercept; while the learning model shows a significant role for learning and a reduced, though still substantial, role for initial productivity and style. Perhaps unsurprisingly, the contributions of time controls are significantly reduced when the learning dynamics are specified.

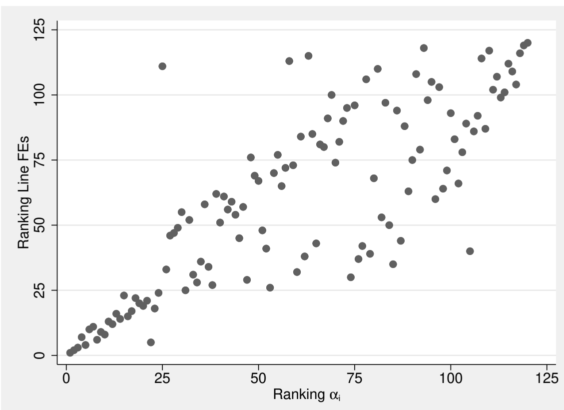

FIGURE 6A
$\hat{\alpha}_i$ Ranking vs $\hat{\hat{\alpha}}_i$ Ranking

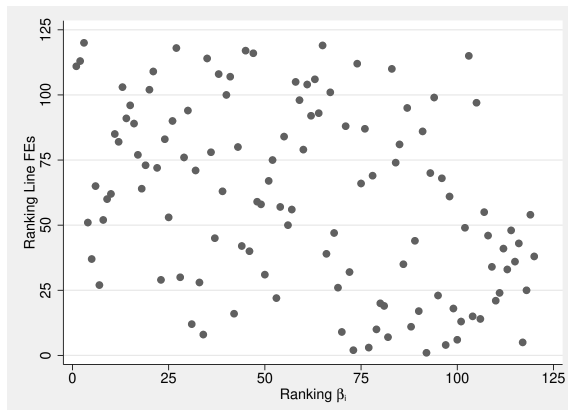

FIGURE 6B
$\hat{\beta}_i$ Ranking vs $\hat{\hat{\alpha}}_i$ Ranking

Note: Figures 6A and 6B show the relationships between the rank of each line in terms of $\hat{\alpha}_i$ and $\hat{\beta}_i$, respectively, against the rank in terms of $\hat{\hat{\alpha}}_i$.

5.2. *Second Stage: managerial quality measures and factors*

In this section, we report and discuss the estimation results from the latent factor measurement system. Remember from the discussion in section 2.3 that we map the complete set of measures from the different modules of the survey using exploratory factor analysis into the following seven dimensions of managerial quality: Tenure, Demographics, Cognitive Skills, Control, Personality, Autonomy, and Attention.[45] Table 4 reports the set of measures used to proxy each latent factor and the estimated loading for each. To establish the informativeness of each measure, we compute

44. We do the analogous exercise to compare $\hat{\hat{\alpha}}_i$, $\hat{\alpha}_i$, and $\hat{\beta}_i$ to supervisor pay in Figure A5 through Figure A7 in the Appendix. No discernible pattern appears, consistent with the limited pass-through of productivity to pay depicted in the simulations below and with the limited degree of pay for performance discussed in section 2 above.

45. The details of the variable construction and exploratory factor analysis are presented in Appendix B.

the signal content (i.e., the variance of the contribution to the latent factor over the residual variance of the measure). Remember that for each factor we normalized the highest loading measure to a loading of 1 such that the loadings of all other measures are relative to that highest loading measure.

TABLE 4
Loadings and Signals

Measures	Tenure	Demographics	Cognitive Skills	Control	Personality	Autonomy	Attention	Signal
Tenure Supervising Current Line	1							0.327
Tenure as Supervisor	0.838							0.517
Tenure in Garment Industry	0.598							0.195
Total Years Working	0.389							0.079
Egalitarianism		1						0.996
Demographic Similarity		0.023						0.005
Arithmetic			1					0.503
Digit Span Recall			0.797					0.344
Arithmetic Correct (%)			0.263					0.204
Internal Locus of Control				1				0.348
Patience				0.329				0.029
Risk Aversion				0.089				0.002
Perseverance					1			0.783
Conscientiousness					0.992			0.746
Self-Esteem					0.958			0.684
Psychological Distress					-0.403			0.101
Initiating Structure						1		0.927
Consideration						0.866		0.792
Self-Assessment						0.085		0.013
Autonomous Problem-Solving						0.062		0.005
Identifying Production Problems						-0.033		0.003
Monitoring Frequency							1	0.471
Active Personnel Management							0.755	0.248
Issues Motivating Workers, Resistance							0.029	0.002
Efforts to Meet Targets							0.016	0.003
Lack of Communication							-0.181	0.023

Note: The first loading of each factor is normalized to 1. Signal of measure j of factor k is $s_j^k = \frac{\left(\lambda_{j,k}\right)^2 Var\left(\ln\theta_k\right)}{\left(\lambda_{j,k}\right)^2 Var\left(\ln\theta_k\right)+Var\left(\varepsilon_{j,k}\right)}$. The measures were standardized across all supervisors who were surveyed. Learning parameters (α and β) and the mean of log pay (including both monthly salary and production bonus) from November 2014 across supervisors of a line are all included in the extended system but measured with no error, i.e., the corresponding factor loadings are set equal to 1 but omitted from this table.

Table 4 shows that the most informative measure for Tenure is years as a supervisor with a signal of nearly 52%. Second and third most informative is years supervising current line and years in the garment industry with signals of 33% and just under 20%, respectively.[46] Total years working is less informative than the more job and industry-specific measures, with a signal of only 8%. For Demographics, the loading is largest for the measure of egalitarianism with a signal of nearly 100%; while the signal from demographic similarity is effectively 0. It is intuitive that if egalitarianism dominates the factor, demographic similarity will be irrelevant.[47]

For Cognitive Skills, Table 4 shows that all three measures are all informative, although the signal is a bit higher for arithmetic (50%) than for the memory measure (34%). The second

46. Note that tenure supervising current line might reflect relationship between the supervisor and the specific workers on the line or worker composition more generally. We account for this by projecting off the full set of worker fixed effects from both log of line-day efficiency and log of experience stock before conducting the structural estimation. The inclusion of these worker composition controls yields a reduction in the informative signal of this measure for the Tenure factor from over 50%, the highest signal among the set, to 33% and second most informative among the set.

47. It is interesting to note that the inclusion of controls for worker composition changed the orientation of this factor from representing primarily demographic similarity (though with substantial noise) to entirely egalitarianism. That is, it seems once the specific worker identities were accounted for, the contribution of the attitudes of the supervisors with respect to equality were able to more clearly be captured.

arithmetic measure contributes some signal (20%), but mostly serves to purge the others of measurement error. With respect to Control, internal locus of control has the highest loading and a signal of 35% justifying our naming this factor after this measure. Patience and risk aversion also contribute with loadings of .33 and .09, but both contain much more noise with signals of only 3% and .2%, respectively. These measures once again help to eliminate error from the dominant measure of internal locus of control. With respect to Personality, perseverance, conscientiousness, and self-esteem are all highly informative. The three measures present signal of 78%, 75%, and 68%, respectively, and all have loadings near 1, reflecting the strong correlations between them. Psychological distress is less informative than the other three with a loading of -0.4 and a signal of 10%. Note that a higher score on the Kessler scale corresponds to more distress, so a negative loading is what we would expect.

For Autonomy, the two leadership behavior measures, initiating structure and consideration, are highly informative with loadings of 1 and .8 and signals of 93% and 79%, respectively. Autonomous problem-solving, problem identification, and self-assessment contribute little incrementally and are much noisier with signals of only .5%, .3% and 1.3%, respectively, but help to isolate the informative content in the leadership behavior measures. Finally, for Attention, monitoring frequency and active personnel management are the strongest contributors, with monitoring frequency having the strongest loading and active personnel management contributing .76, relatively. The signals of these two dominant measures (47% and 25%, respectively) reflect the importance of having other measures to purge them of noise. The other three measures serve this purpose, contributing little incremental signal of their own.[48]

5.3. *Third Stage: productivity contributions of managerial quality*

Table 5 reports the estimates of the CES functions depicted in equation (3.5) for $\iota \in \{\tilde{\alpha}, \alpha, \beta\}$. We report mean coefficients across 200 bootstrap replications of the full three stage procedure as parameter estimates and standard deviations across replications as errors. We begin on the left with $\tilde{\alpha}_i$ from the static model and then show jointly estimated parameters for α_i and β_i from the learning model in the two right columns. Note that the factors are all standardized for ease of interpretation across coefficients.

On the left, the static model reveals Control to be the largest contributor of all the factors (.48) followed by Attention (.28) and Cognitive Skills (.2). Tenure has a smaller but detectable effect (.04). Demographics, Personality, and Autonomy factors show negligible contributions to productivity.

The results are notably different when analyzing the learning model. The initial productivity equation in the middle column of Table 5 once again shows Control to be the largest contributor followed by Attention but their relative contributions are much closer than in the static model (.39 for Control and .31 for Attention). Tenure shows the third strongest contribution to initial productivity, with Cognitive Skills only half as important as it was in the static model. Autonomy now has a detectable effect on (.07), but Demographics and Personality still are negligible. In the right most column, contributions to the rate of learning differ even more markedly. Attention contributes most strongly (.32) followed by Control (.26) and Tenure (.22). Autonomy shows its strongest effects on the rate of learning (.12), nearly twice as strong as its effect on initial productivity; while Cognitive Skills contributes least to the rate of learning, though still detectably (.07).

48. Note that we would expect less communication with workers and upper management regarding production to indicate less managerial attention or effort, so a negative loading is intuitive. We might expect the same for issues motivating workers, but find a loading of effectively 0.

TABLE 5
Contributions of Managerial Quality to Productivity

	Static Model ($\tilde{\alpha}$)	Initial Productivity (α_i)	Rate of learning (β_i)
Tenure	0.043	0.126	0.222
	(0.024)	(0.049)	(0.041)
Demographics	0.002	0.002	0.0001
	(0.001)	(0.001)	(0.0006)
Cognitive Skills	0.199	0.108	0.069
	(0.041)	(0.077)	(0.082)
Control	0.478	0.391	0.266
	(0.009)	(0.058)	(0.042)
Personality	0.000	0.0000	0.0001
	(0.000)	(0.0002)	(0.0002)
Autonomy	0.003	0.068	0.118
	(0.009)	(0.034)	(0.026)
Attention	0.277	0.305	0.323
	(0.021)	(0.061)	(0.060)
Productivity Parameter	1.006	1.006	1.011
	(0.004)	(0.001)	(0.002)
Complementarity Parameter	-0.505	-0.227	0.228
	(0.034)	(0.075)	(0.052)

Note: Results reflect 200 bootstrap replications with mean coefficients across replications reported as parameter estimates and standard deviations across replications reported as errors in parentheses. The static model on the left is a single equation model; while the two columns on the right are the results of a jointly estimated system of nonlinear equations. We bootstrap the entire procedure – all three stages – from the top. To construct the bootstraps standard errors, we construct residuals for the first stage, and randomly resample the residuals, stratifying by manager-style pair to preserve the match structure of the observations. We then re-estimate the line and style fixed effects, and estimate the second and third stage. We repeat this procedure 200 times, and use the means and standard deviations of estimates across replications for inference.

Taken together, the results in Table 5 highlight that, though some dimensions like Attention and Control contribute strongly to both initial productivity and rates of learning, other dimensions of managerial quality such as Autonomy and Tenure impact productivity most strongly by enabling faster learning-by-doing of production lines and therefore are underemphasized or even altogether ignored when studying a static model of productivity. On the other hand, dimensions like Cognitive Skills contribute more to initial productivity and therefore would be overemphasized if dynamics are ignored.[49]

That managerial practices illustrating greater attention to production issues and autonomy in implementing changes would be important for rapid learning is quite consistent with our understanding of how supervisors enable learning in this context. That is, the main ways in which production line supervisors can improve the productivity of their lines over the life of a production run are to monitor for machine calibration issues and bottle necks, reorganize

49. In order to assess how much of the overall variation in initial productivity/learning-by-doing is explained by the factors, a traditional R^2 cannot be easily calculated because the model is a system of nonlinear equations. Instead, we calculate the implied R^2 for the full system of nonlinear equations by vectorizing the system, and using the vectorized residuals to compute the R^2. The R^2 computed in this way is 0.89, indicating a large proportion of the variation in the initial productivity/learning-by-doing is explained by the factors with the imposed structure.

the sequence of operations, and adjust allocations of workers to machine operations to relieve production imbalances. This process improvement underlying learning in manufacturing settings has been described in related work like Levitt et al. (2013). It also stands to reason that a supervisor with more experience making these tweaks would be more effective as indicated by the role of Tenure.

We also note that overall the two most important dimension of managerial quality appear to be Attention and Control, both more impactful than traditionally emphasized dimensions like Cognitive Skills and Tenure. These two dimensions have received relatively little emphasis in the literature to date. Recent studies have begun to model and document the impacts of managerial inattention on firm performance (Bandiera et al., 2014; Halac and Prat, 2016; Hortaçsu et al., 2017), but our results are among the first to emphasize the importance of internal locus of control. Notably, the Personality factor, reflecting more commonly studied traits like conscientiousness and perseverance, has no detectable contribution of its own after accounting for these other dimensions; neither does the Demographics factor which captures discriminatory preferences like those studied in Hjort (2014).

However, it should be noted that the factors are allowed to be, and in fact appear to be in Table A4 in the Appendix, correlated with each other such that screening on Personality, though not seemingly productivity-enhancing in its own right, may still deliver gains in productivity by way of correlated dimensions of managerial quality. This estimated correlation validates the specific latent factor measurement system imposed in the second step, as more standard latent factor models would impose orthogonality. We next explore the implications of this correlation structure along with the composite implications of distinct factor contributions across initial productivity and rates of learning in simulations.

The simulations will also make use of the nonlinearity between factors allowed for in the structural estimation. Estimates in Table 5 show imperfect but significant complementarity between the factors in determining initial productivity (paralleled and amplified in the static model), but imperfect substitutability among the factors in determining the rate of learning. These estimates validate, in addition to allowing for correlation among the latent factors, the added flexibility allowed by the CES form, albeit a compromise, as linear combinations of the factors would not fit the data as well or yield the same counterfactual insights.

As discussed in section 4.1.5 above, we also perform the covariance shrinkage method proposed by Best et al. (2019), to account for the correlation between the estimation errors of the two vectors of fixed effects that might arise due to limited mobility bias. The results are reported in Tables A9 and A10 in the Appendix and closely mirror the main results presented here.

5.4. *Counterfactual Simulations*

5.4.1. Productivity. In this section, we simulate the contribution to productivity of a one standard deviation (SD) increase in each of the dimensions of quality to leverage the full complexity and flexibility of the structural estimation. Specifically, we substitute the estimated functions for α_i and β_i presented in Table 5 into the first stage (equation 3.1) and compute the impact of an increased stock of each factor (as estimated in the second stage) on productivity, weighting the contributions of α_i and β_i according to how they relate to line-level average productivity in the raw data.[50] We first evaluate productivity with each factor in each learning

50. Note that, armed with the estimated joint distribution of α_i and β_i (as well as the managerial quality factors), one way to conduct the simulation might be to increase each factor by one SD in the estimated CES function for either α_i and β_i and let the other parameter be determined solely by the estimated joint distribution. This would obviate the need for any weighting when combining α_i and β_i to calculate productivity. However, this would involve deciding which of the

parameter fixed to its mean (baseline), and then reevaluate with each factor increased sequentially by one SD.

We use the estimated covariance structure of the factors in the population and compute the impact of an increase of factor i by κ_i, i.e., $E(\ln\boldsymbol{\theta}|\ln\theta_i = \kappa_i)$ where $\kappa_i = \sqrt{\sigma_{ii}}$ and $\sigma_{ii} = var(\theta_i)$. The computation of $E(\ln\boldsymbol{\theta}|\ln\theta_i = \kappa_i)$ depends on the nature of the multivariate distribution recovered for $\ln\boldsymbol{\theta}$, thus

$$E(\ln\boldsymbol{\theta}|\ln\theta_i = \kappa_i) = (\sigma_{1i}/\sigma_{ii}, \cdots, \sigma_{Ki}/\sigma_{ii})'\kappa_i$$

where $\sigma_{ij} = var(\theta_i, \theta_j)$. This procedure is similar to the generalized impulse response functions proposed in the time series context by Pesaran and Shin (1998).[51]

We present the correlation structure between factors used in these screening simulations in Table A4 in the Appendix. Control and Personality, and to some degree Cognitive Skills, are most correlated with other factors, while Attention and to some degree Autonomy and Tenure exhibit weaker correlations. This nuanced pattern confirms that the factors are not randomly assigned in the population. As such, interpretation of these counterfactual simulations relies crucially on this structurally estimated correlation.

We evaluate simulated productivity gains under three alternate scenarios. First, we assume the experience stock as given by equation (3.2) on the first day of the order and the length of the order are their mean values observed in the data. We simulate productivity along this order cycle according to equation (3.1) and report the mean across simulated days. Then, we assume two other order patterns: one in which a line (or factory) alternates producing only two different styles each for 5 days, and one in which the alternating orders across the 2 styles are for 30 days each. We evaluate the productivity over a simulated 300 days (roughly a year of production) under each of these two scenarios (i.e., 60 orders in the short order scenario and 10 orders in the long) and report the mean of the line-day productivity that prevails in each case. The latter two exercises help to investigate how a factory would value dimensions of managerial quality differently depending on the scope for learning in the orders they tend to produce. Note many small scale domestic producers in India and other similar developing countries with large garment industries indeed specialize in smaller orders; while export suppliers to multinational brands tend to produce disproportionately large volume orders.

When presenting the results in Table 6, we group the factors to guide interpretation. Some dimensions like Tenure and Demographics are traditionally valued and more easily and often observed at the time of hiring. Others are less readily observed like Cognitive Skills, Control, and Personality, especially at lower levels of management in labor-intensive manufacturing contexts like ours. On the other hand, dimensions like Attention and Autonomy representing management practices and behaviors are underemphasized but potentially trainable.

Several interesting insights are revealed by the simulations. First, some of the less readily screened dimensions of quality like Control and Cognitive Skills are the most impactful, as compared to more commonly screened dimensions such as Tenure. Indeed, the simulations confirm that the most impactful dimension is Control which is a novel trait for both the academic literature and industry practice; and next is Attention for which empirical evidence is only

estimated functions (α_i or β_i) to prioritize. As we believe each function contributes informative value to the simulation, we have chosen to calculate the impact of increasing each factor by one SD in both equations. Then we use the relationship between efficiency and α_i and β_i in the raw data (as in the first stage) to weight the simulated impact on α_i and β_i to reproduce the resulting overall impact on productivity. We obtain the weights on α_i and β_i by regressing the average efficiency of each production line on α_i and β_i. The weights on α_i and β_i we use for the simulations are 1.042 and 0.518, respectively.

51. See also Pesaran (2015).

recently starting to build. Second, though Demographics and Personality appeared negligible in the structural estimates of equation (3.5) for $\iota \in \{\alpha, \beta\}$ reported in Table 5, the simulations show a substantial gain in productivity from screening on these dimensions, reflecting the correlation across factors recovered in the second stage and presented in Table A4 in the Appendix.

TABLE 6
Simulated Contributions to Productivity

Factor	Mean	Short orders	Long orders
Screening (Easily Observed)			
Tenure	0.240	0.172	0.347
	(0.043)	(0.034)	(0.059)
Demographics	0.168	0.132	0.226
	(0.014)	(0.010)	(0.019)
Screening (Costly to Observe)			
Cognitive Skills	0.292	0.229	0.391
	(0.055)	(0.040)	(0.078)
Control	0.426	0.335	0.571
	(0.057)	(0.043)	(0.079)
Personality	0.237	0.178	0.331
	(0.023)	(0.023)	(0.041)
Training			
Autonomy	0.166	0.119	0.239
	(0.029)	(0.022)	(0.040)
Attention	0.303	0.227	0.425
	(0.019)	(0.015)	(0.025)

Note: Table 6 shows the contribution to productivity of an increase of each factor by one standard deviation. We use the covariance structure of the factors to compute the impact on productivity. We first assume the experience stock as given by equation (3.2) on the first day of the order and the length of the order are their mean values observed in the data. Then we assume an order pattern in which a line alternated producing two styles each for 5 days (short orders) and for 30 days (long orders). We evaluate productivity over a simulated 300 days under these two scenarios.

Finally, note that several dimensions contribute substantially more to productivity under the long order profile than under short orders; while others contribute only modestly more when the scope for learning is greater. Figure 7 makes this pattern more explicit by graphing for each factor the ratio of simulated productivity gains for the long order profile over the short. Note that the value of all factors is increasing in the scope for learning, but Tenure and Autonomy and to a somewhat lesser degree Attention and Personality all appear to contribute more strongly through learning than do the others. That is, Cognitive Skills, Control, and, by way of correlations with other factors, Demographics should seemingly be emphasized irrespective of the length of the order; while Tenure, Autonomy, Attention and, again by way of correlations with other factors, Personality should be of particular interest to factories producing primarily large volume orders.

Put another way, the priority ranking of dimensions of managerial quality on which the firm should focus in screening and training policies is Control first, followed by Cognitive Skills, Attention, Personality, Tenure, Demographics, and Autonomy last when orders are short. But when orders are longer, even under the mean order pattern observed in the sample, Attention, Tenure, and Autonomy all rise in importance relative to Cognitive Skills, Personality, and Demographics.

Note that these simulations utilize the estimated correlation structure for all dimensions of quality. This is clearly appropriate for any dimension for which the most easily implemented

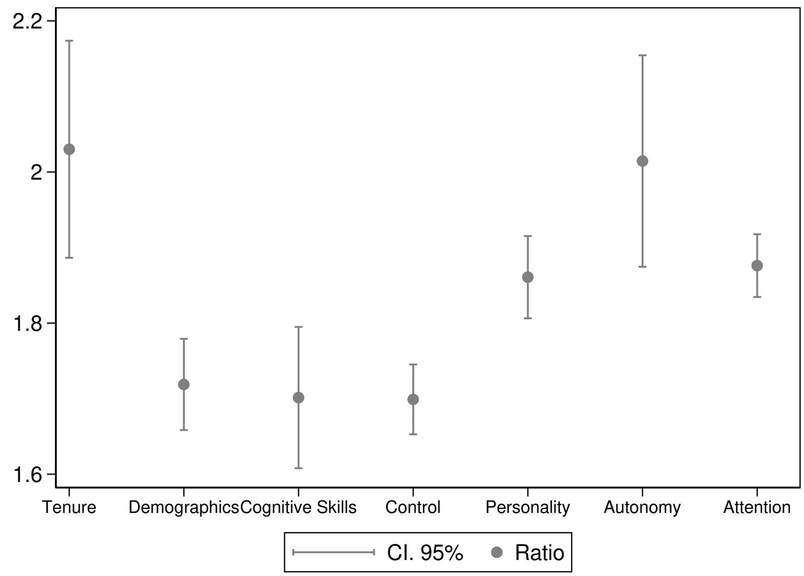

FIGURE 7

Ratio Long/Short Orders

Note: Figures 7 shows the impact on productivity of an increase of each factor by one standard deviation for the long over the short order profile. We use the covariance structure of the factors to compute the impact on productivity. For the long and short profiles we assume an order pattern in which a line alternated producing two styles each for 30 and 5 days, respectively. We evaluate productivity over a simulated 300 days under these two scenarios.

policy would be to screen at the point of hiring. However, for practices and behaviors which we might think are mutable by way of training like Autonomy and Attention, it might also be possible to increase the stock of these dimensions of quality independently. To evaluate this possibility, we repeat the counterfactual simulations for these two factors ignoring their correlations with the other factors. The results reported in Table A7 in the Appendix show roughly 25% larger gains from independent increases in each of these dimensions, and a further amplified value of Autonomy under long orders. It is not clear that a focused training could in fact increase each factor without changing other dimensions of skill, but these results indicate that such a training would be even more impactful if possible.

Of course, the decisions of which policy – screening or training – and which dimensions to prioritize depend also on corresponding impacts on the pay needed to attract and retain these higher quality supervisors. For that, we must conduct an analogous third stage estimation for observed pay of supervisors as well as corresponding simulations, and compare results with these productivity simulations.

5.4.2. Supervisor Pay. Having estimated the contributions of the seven latent factors to the learning parameters and simulated impacts of skill increases on composite productivity, we next test if there exists a relationship between these seven factors and supervisor pay observed in the sample. If pay reflects the marginal productivity of labor, as a standard model of a perfectly competitive labor market would predict, we may expect similar results to the ones presented in Table 5. However, imperfect information on the part of the employer (or competing employers) regarding quality of the managers, particularly less easily measured or observed dimensions of quality, may lead the firm to rely just on the observable characteristics, like tenure or demographics to determine the pay scheme (or only force the firm to reward these observable dimensions).

We interpret this exercise as a partial equilibrium assessment on the margin. That is, we assume that the supply in the market of supervisors of each skill profile is not impacted by the simulated screening or training, nor is the market price for each dimension of skill.[52] We simply intend to evaluate using analogous procedures to those employed above for productivity whether the in-sample pattern of pay across skill profiles of supervisors reflects proportionately the productive value of each dimension we have estimated.[53]

To test the link between the seven latent factors and supervisor pay, we follow the same approach as we did for productivity. We use data on salary paid by the firm to each of the managers during the month of the survey, November 2014, and include the monetary bonuses that are associated with the productivity of the lines. Remember that we included the log of this pay measure in the measurement system in the second stage of our empirical strategy. Accordingly, we can draw synthetic datasets from the joint distribution of factors and supervisor pay just as we did for the learning parameter analysis above. We estimate an analogue to equation (3.5) with log of supervisor pay as the outcome.

Then, to best assess the relative pass-through of productivity contributions of factors to pay, we perform analogous simulations for pay to the productivity simulations summarized in Table 6, and compare results across pay and productivity simulations for each factor. We focus only on the mean experience profile and for these simulations. Finally, we compute the pass-through of productivity to pay by dividing the simulated change in pay by the simulated change in productivity for the one SD increase in each factor.

Table 7 presents the results and is structured similarly to Table 6 in terms of grouping of factors. We see that Control exhibits the smallest pass-through to pay followed by Cognitive Skills as measured by memory and arithmetic. Both represent traits and skills which are less commonly screened during hiring or even over the course of employment. Autonomy and Tenure exhibit the largest pass-through to pay. It is unsurprising that Tenure is among the most appropriately valued in pay as it is perhaps the most commonly observed and emphasized dimension of skill across occupations, industries, and countries. Autonomy is not as readily measured in most contexts but represents a leadership style which might intuitively be valued even if casually observed by upper management.

Note that we focus our interpretation on the relative values of pass-through across dimensions rather than the absolute values. Overall, pass-through is quite low, ranging from 66% to as little as 39%. This is consistent with the firm paying almost entirely fixed salaries with a limited role for performance-contingent bonuses as indicated by the summary statistics on pay.[54] However, there are many reasons for less than full pass-through. The firm may exhibit some monopsonistic power in the local market or marginal revenue product may differ from the marginal productivity of the line we analyze in these simulations. But these market or firm level features are less likely to be reflected in relative comparisons of pass-through across dimensions of skill. We interpret

52. Though our data is from the largest export producer of garments in India, they are still relatively small compared to the entire size of the market for garment sector employees. The Indian garment industry employs nearly 13 million workers in factories like the ones we study, whereas this firm employs just over 100,000.

53. Note we also implicitly assume that the marginal productivity of the line well-approximates or is sufficiently proportional to the marginal revenue product of the firm. This assumption is based on data from and conversations with the accounting department confirming that labor productivity of the line is the second largest cost category after materials like textiles which are used only proportionately to labor productivity, such that gains in efficiency can be directly translated into revenue and profit.

54. This pattern is also consistent with evidence from a previous study in a similar context (Bloom et al., 2013).

TABLE 7
Simulated Pass-through to Pay

	Contribution to Productivity	Contribution to Pay	Pass-through
	Screening (Easily Observed)		
Tenure	0.240	0.154	64.77%
	(0.043)	(0.018)	
Demographics	0.168	0.075	44.75%
	(0.014)	(0.008)	
	Screening (Costly to Observe)		
Cognitive Skills	0.292	0.121	41.26%
	(0.055)	(0.025)	
Control	0.426	0.164	38.76%
	(0.057)	(0.027)	
Personality	0.237	0.131	55.48%
	(0.030)	(0.014)	
	Training		
Autonomy	0.166	0.109	65.62%
	(0.029)	(0.013)	
Attention	0.303	0.164	54.03%
	(0.019)	(0.009)	

Note: Table 7 shows the contributions (percentage change) to productivity and pay of an increase of each factor by one standard deviation. The associated changes in all other factors as given by the covariance structure among factors. We first assume the experience stock as given by equation (3.2) on the first day of the order and the length of the order are their mean values observed in the data. We compute the pass-through of productivity to pay, dividing the contribution to pay by the contribution to productivity. Standard errors in parentheses are based on 200 bootstrap replications.

these patterns as more likely reflective of the ability of supervisors to signal these qualities on the market or the degree to which firms measure or value these dimensions. The variation in pass-through across factors is evidence that the firm does perceive and reward in salary, despite limited performance pay, some dimensions of skill more than others. Taken together, the results in Tables 6 and 7 indicate that the firm could perhaps benefit most from screening on Control and training in Attention without likely having to increase pay as commensurately.

6. CONCLUSION

Using granular productivity data from 6 factories of one of the largest garment exporters in the world, we document substantial variation in productivity within the firm both across production lines and within lines over the course of an order. Matching these data to a comprehensive survey of managerial skills, traits, and practices across supervisors of 120 production lines, we show that multiple dimensions of managerial quality appear to coincide with these productivity fluctuations in nuanced ways. Some dimensions of quality are associated with higher minimum or initial productivity and others with higher maximum or peak productivity or both. This combination of administrative data on productivity across teams within the firm and comprehensive survey measures of managerial quality across the supervisors of these different teams is rare.

We leverage this novel data to estimate a dynamic model of line-day productivity to recover line-specific parameters for initial productivity and rate of learning. Building on an AKM-style two-way fixed effects model, we account for style or order assignments across lines as well as variation in worker composition to identify the line-specific contributions to productivity. We then estimate a latent factor measurement system to extract seven distinct dimensions of managerial quality from the universe of noisy and potentially redundant survey measures available, allowing for correlation between the recovered factors. Finally, we assume a CES form by which these factors are combined to produce the line-specific parameters from the first stage and ultimately the observed line-day productivity in the data.

This structural procedure allows us to evaluate counterfactual simulations for productivity under differing stocks of managerial skill and differing order lengths. We also leverage the in-sample pattern between line productivity and supervisor pay to identify opportunities for screening on or training in particularly productive dimensions of managerial skill with relatively low impact on labor costs. Taken together, our results emphasize the value of some novel traits such as Control, reflecting the supervisor's belief in the ability to affect change rather than acquiesce to chance or predetermination, even above traditional dimensions like Tenure. They also highlight the importance of practices or behaviors such as Attention, measured primarily by frequency of monitoring and effort invested in personnel management. Both are revealed to be opportunities for gains, yielding large productivity gains with relatively low pass-through to pay.

The results of these simulations are remarkably consistent with findings from a randomized training trial in similar factories. Adhvaryu et al. (2021) find that a managerial skills training for production line supervisors increased practices and behaviors related to Autonomy and Attention by roughly .1 SD. Line productivity increased as a result of training by roughly 3 pp from a base of roughly 55% efficiency, or more than 5% from the mean. The results of the simulations in this study indicate that a 1 SD increase in Autonomy or Attention would increase productivity by 20 to 36%. Accordingly, we interpret the results of the experiment as a reasonable validation of the simulations here. Furthermore, the training only yielded an increase in salary of less than 1%, consistent with the low pass-through depicted by the simulations. Nevertheless, further work is still needed to validate predicted gains from screening on other dimensions of skill such as Control and to investigate with more nuance the ways in which these multiple dimensions of skill interact with each other to generate productivity.

Aknowledgment Thanks to Nick Bloom, Hashem Pesaran, Raffaella Sadun, John Van Reenen, and seminar participants at the NBER, Columbia, Northwestern Kellogg, Harvard, McGill, Michigan, USC, HEC Paris, Clark, Stockholm University, and the MIT/Harvard/Stanford Meeting on Empirical Management for helpful comments. Jorge Tamayo gratefully acknowledges funding from the USC Dornsife INET graduate student fellowship. All errors are our own.

Data Availability Statement

The data and code underlying this research is available on Zenodo at https://doi.org/10.5281/zenodo.6856836.

Bibliography

Abowd, J. M., R. H. Creecy, and F. Kramarz (2002). Computing person and firm effects using linked longitudinal employer-employee data. Technical report.

Abowd, J. M., R. H. Creecy, F. Kramarz, et al. (2002). Computing person and firm effects using linked longitudinal employer-employee data. Technical report, Center for Economic Studies, US Census Bureau.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*(2), 251–334.

Abowd, J. M., P. Lengermann, and S. Pérez-Duarte (2004). Are good workers employed by good firms? a simple test of positive assortative matching models. Technical report.

Acemoglu, D., P. Aghion, C. Lelarge, J. Van Reenen, and F. Zilibotti (2007). Technology, information, and the decentralization of the firm. *The Quarterly Journal of Economics 122*(4), 1759–1799.

Adhvaryu, A., V. Bassi, A. Nyshadham, and J. Tamayo (2021). No line left behind: Assortative matching inside the firm. Technical report, Working Paper.

Adhvaryu, A., J.-F. Gauthier, A. Nyshadham, and J. Tamayo (2021). Absenteeism, productivity, and relational contracts inside the firm. Technical report, Working Paper.

Adhvaryu, A., N. Kala, and A. Nyshadham (2018). The skills to pay the bills: Returns to on-the-job soft skills training. Technical report, National Bureau of Economic Research.

Adhvaryu, A., N. Kala, and A. Nyshadham (2019). Management and shocks to worker productivity. Technical report, National Bureau of Economic Research.

Adhvaryu, A., E. Murathanoglu, and A. Nyshadham (2021). (mis)allocation of managerial training inside the firm. Technical report, Working Paper.

Aghion, P., N. Bloom, B. Lucking, R. Sadun, and J. Van Reenen (2017). Turbulence, firm decentralization and growth in bad times. *working paper*.

Alfonsi, L., O. Bandiera, V. Bassi, R. Burgess, I. Rasul, M. Sulaiman, A. Vitali, et al. (2017). Tackling youth unemployment: Evidence from a labor market experiment in uganda. *STICERD-Development Economics Papers*.

Almlund, M., A. L. Duckworth, J. Heckman, and T. Kautz (2011). Personality psychology and economics. In *Handbook of the Economics of Education*, Volume 4, pp. 1–181. Elsevier.

Alvarez, J., F. Benguria, N. Engbom, and C. Moser (2018). Firms and the decline in earnings inequality in brazil. *American Economic Journal: Macroeconomics 10*(1), 149–89.

Andrews, M. J., L. Gill, T. Schank, and R. Upward (2008). High wage workers and low wage firms: negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society: Series A (Statistics in Society) 171*(3), 673–697.

Andrews, M. J., L. Gill, T. Schank, and R. Upward (2012). High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias. *Economics Letters 117*(3), 824–827.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2015). Estimating the production function for human capital: Results from a randomized control trial in colombia. Technical report, National Bureau of Economic Research.

Attanasio, O., C. Meghir, and E. Nix (2015). Human capital development and parental investment in india. Technical report, National Bureau of Economic Research.

Bandiera, O., A. Prat, S. Hansen, and R. Sadun (2020). Ceo behavior and firm performance. *Journal of Political Economy 128*(4), 1325–1369.

Bandiera, O., A. Prat, R. Sadun, and J. Wulf (2014). Span of control and span of attention. *Harvard Business School strategy unit working paper* (12-053), 14–22.

Bassi, V. and A. Nansamba (2017). Information frictions in the labor market: Evidence from a field experiment in uganda.

Benkard, C. L. (2000). Learning and forgetting: The dynamics of aircraft production. *The American Economic Review 90*(4), 1034–1054.

Bertrand, M. and E. Duflo (2017). Field experiments on discrimination. In *Handbook of economic field experiments*, Volume 1, pp. 309–393. Elsevier.

Bertrand, M. and A. Schoar (2003). Managing with style: The effect of managers on firm policies. *The Quarterly Journal of Economics 118*(4), 1169–1208.

Best, M. C., J. Hjort, and D. Szakonyi (2019). Individuals and organizations as sources of state effectiveness. Technical report, National Bureau of Economic Research.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013). Does management matter? evidence from india. *The Quarterly Journal of Economics 1*(51), 51.

Bloom, N., L. Garicano, R. Sadun, and J. Van Reenen (2014). The distinct effects of information technology and communication technology on firm organization. *Management Science 60*(12), 2859–2885.

Bloom, N., R. Lemos, R. Sadun, D. Scur, and J. Van Reenen (2014). Jeea-fbbva lecture 2013: The new empirical economics of management. *Journal of the European Economic Association 12*(4), 835–876.

Bloom, N., R. Sadun, and J. Van Reenen (2016). Management as a technology? Technical report, National Bureau of Economic Research.

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

Bloom, N. and J. Van Reenen (2010). Why do management practices differ across firms and countries? *The Journal of Economic Perspectives*, 203–224.

Bloom, N. and J. Van Reenen (2011). Human resource management and productivity. *Handbook of labor economics 4*, 1697–1767.

Boissiere, M., J. B. Knight, and R. H. Sabot (1985). Earnings, schooling, ability, and cognitive skills. *The American Economic Review 75*(5), 1016–1030.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. Ter Weel (2008). The economics and psychology of personality traits. *Journal of human Resources 43*(4), 972–1059.

Bowles, S., H. Gintis, and M. Osborne (2001). The determinants of earnings: A behavioral approach. *Journal of economic literature 39*(4), 1137–1176.

Bresnahan, T. F., E. Brynjolfsson, and L. M. Hitt (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The Quarterly Journal of Economics 117*(1), 339–376.

Burks, S. V., B. Cowgill, M. Hoffman, and M. Housman (2015). The value of hiring through employee referrals. *The Quarterly Journal of Economics 130*(2), 805–839.

Card, D., A. R. Cardoso, and P. Kline (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics 131*(2), 633–686.

Card, D., J. Heining, and P. Kline (2013). Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly journal of economics 128*(3), 967–1015.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review 104*(9), 2593–2632.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics 132*(4), 1593–1640.

Donato, K., G. Miller, M. Mohanan, Y. Truskinovsky, and M. Vera-Hernandez (2017). Personality traits and performance contracts: Evidence from a field experiment among maternity care providers in india. *American Economic Review 107*(5), 506–10.

Dustmann, C., A. Glitz, U. Schönberg, and H. Brücker (2015). Referral-based job search networks. *The Review of Economic Studies 83*(2), 514–546.

Ellison, S. F., C. Snyder, and H. Zhang (2018). Costs of managerial attention and activity as a source of sticky prices: Structural estimates from an online market. Technical report, National Bureau of Economic Research.

Ellison, S. F. and C. M. Snyder (2014). An empirical study of pricing strategies in an online market with high-frequency price information.

Frederiksen, A., L. B. Kahn, and F. Lange (2020). Supervisors and performance management systems. *Journal of Political Economy 128*(6), 2123–2187.

Gibbons, R. and M. Waldman (2004). Task-specific human capital. *The American Economic Review 94*(2), 203–207.

Groves, T., Y. Hong, J. McMillan, and B. Naughton (1994). Autonomy and incentives in chinese state enterprises. *The Quarterly Journal of Economics 109*(1), 183–209.

Halac, M. and A. Prat (2016). Managerial attention and worker performance. *American Economic Review 106*(10), 3104–32.

Heckman, J. J. and T. Kautz (2012). Hard evidence on soft skills. *Labour economics 19*(4), 451–464.

Hjort, J. (2014). Ethnic divisions and production in firms. *The Quarterly Journal of Economics 129*(4), 1899–1946.

Hoffman, M., L. B. Kahn, and D. Li (2017). Discretion in hiring. *The Quarterly Journal of Economics 133*(2), 765–800.

Hoffman, M. and S. Tadelis (2018). People management skills, employee attrition, and manager rewards: An empirical analysis. Technical report, National Bureau of Economic Research.

Hortaçsu, A., F. Luco, S. L. Puller, and D. Zhu (2017). Does strategic ability affect efficiency? evidence from electricity markets. Technical report, National Bureau of Economic Research.

Jovanovic, B. (1979). Firm-specific capital and turnover. *Journal of political economy 87*(6), 1246–1260.

Judge, T. A., R. F. Piccolo, and R. Ilies (2004). The forgotten ones? the validity of consideration and initiating structure in leadership research. *Journal of applied psychology 89*(1), 36.

Kahn, L. B. (2013). Asymmetric information between employers. *American Economic Journal: Applied Economics 5*(4), 165–205.

Kahn, L. B. and F. Lange (2014). Employer learning, productivity, and the earnings distribution: Evidence from performance measures. *The Review of Economic Studies 81*(4), 1575–1613.

Kane, T. J. and D. O. Staiger (2008). Estimating teacher impacts on student achievement: An experimental evaluation. Technical report, National Bureau of Economic Research.

Korman, A. K. (1966). "consideration","initiating structure", and organizational criteria-a review. *Personnel Psychology 19*(4), 349–361.

Lange, F. (2007). The speed of employer learning. *Journal of Labor Economics 25*(1), 1–35.

Lazear, E. P., K. L. Shaw, and C. T. Stanton (2015). The value of bosses. *Journal of Labor Economics 33*(4), 823–861.

Levitt, S. D., J. A. List, and C. Syverson (2013). Toward an understanding of learning by doing: Evidence from an automobile assembly plant. *Journal of Political Economy 121*(4), 643–681.

Mincer, J. and H. Ofek (1982). Interrupted work careers: Depreciation and restoration of human capital. *Journal of human resources*, 3–24.

Mincer, J. A. et al. (1974). Schooling, experience, and earnings. *NBER Books*.

Neal, D. (1995). Industry-specific human capital: Evidence from displaced workers. *Journal of labor Economics 13*(4), 653–677.

Pesaran, M. H. (2015). *Time Series and Panel Data Econometrics*. Oxford University Press.

Pesaran, M. H. and Y. Shin (1998). Generalized impulse response analysis in linear multivariate models. *Economics Letters 58*(1), 17–29.

Reis, R. (2006). Inattentive producers. *The Review of Economic Studies 73*(3), 793–821.

Schoar, A. (2014). The importance of being nice: Supervisory skill training in the cambodian garment industry. *working paper*.

Schönberg, U. (2007). Testing for asymmetric employer learning. *Journal of Labor Economics 25*(4), 651–691.

Shahi Exports Pvt Ltd (2018). Production lines and supervisors. Unpublished Data Set, (accessed January 2018).

Stogdill, R. M. and A. E. Coons (1957). Leader behavior: Its description and measurement.

Thompson, P. (2001). How much did the liberty shipbuilders learn? new evidence for an old case study. *Journal of Political Economy 109*(1), 103–137.

Thompson, P. (2012). The relationship between unit cost and cumulative quantity and the evidence for organizational learning-by-doing. *Journal of Economic Perspectives 26*(3), 203–224.

Topel, R. (1991). Specific capital, mobility, and wages: Wages rise with job seniority. *Journal of political Economy 99*(1), 145–176.