HARVARD | BUSINESS | SCHOOL

DOCTORAL PROGRAMS

DISSERTATION ACCEPTANCE CERTIFICATE

(To be placed in First Library Copy)

The undersigned, approved by the Chair of the Doctoral Programs, have examined a dissertation entitled

Why Do Firm Practices Differ? Examining the Selection and Implementation of Organizational Practices

presented by Megan Lynn Lawrence

candidate for the degree of Doctor of Business Administration and hereby certify that it is worthy of acceptance.

Signature ....

Felix Oberholzer-Gee, Chair

Signature ....

Jan W. Rivkin

Signature ....

Juan Alcacer

Signature ....

Victor Manuel Bennett

Date 3/10/16

WYSS | SOLDIERS FIELD | BOSTON, MA 02163 | Ph 617.495.6101 | Fx 617.496.2203 | doctoralprograms@hbs.edu | GEORGE F. BAKER FOUNDATION

Why Do Firm Practices Differ?

Examining the Selection and Implementation of Organizational Practices

A dissertation presented

by

Megan Lynn Lawrence

to

the Strategy Unit at Harvard Business School

in partial fulfillment of the requirements

for the degree of

Doctor of Business Administration

in the subject of

Strategy

Harvard University

Cambridge, Massachusetts

March 2016

Dissertation Advisor:
Professor Felix Oberholzer-Gee

Author:
Megan Lynn Lawrence

Why Do Firm Practices Differ?

Examining the Selection and Implementation of Organizational Practices

ABSTRACT

This dissertation is comprised of three studies investigating sources of variation in firm practices. Firm practices may differ both due to differences in the practices firms choose to implement – different types of firms may make different selections – and due to differences in implementation success of similar practices – variation in internal firm conditions may result in differences in otherwise similar practices. The first essay examines a difference in firm practice selection whereas the second and third essays examine differences in firm practice implementation. Essay one considers how ownership impacts the management practices implemented by firms, specifically considering the founder CEO firm's adoption of management practices as compared to firms with other owner-manager types. Founder CEO firms adopt fewer management practices than firms under other ownership structures, both due to a lack of awareness about the lower quality of their practices and due to greater value placed on the nonpecuniary benefit provided by potentially less efficient but power-preserving practices. Essays two and three use data from a Fortune 100 retail chain that implemented a new restocking practice across a subset of its retail stores. Essay two examines how prior experience with the old restocking practice impacts a team's ability to perform and learn the new restocking practice. Teams with greater

exposure to the old practice perform worse at first – due to experiencing a competency trap – but then improve more rapidly – due to greater efficiency of communication and coordination. Essay three focuses on the impact of pilot use when rolling out the new practice, proposing that a main function of pilot implementations is to allow for vicarious learning opportunities for stores subsequently implementing the practice. The relative performance of the pilot stores as well as the contextual similarity of these stores to the stores learning from them matters a great deal. Nonpilot stores increasingly rely only on their own experiences rather than the pilots' experiences in instances where the learning opportunities become less obvious.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

This process of acquiring a doctorate has been by far the craziest endeavor I have ever undertaken, and I would be remiss if I didn't take an opportunity, upon its completion, to thank those people that have supported me along the way.

First and foremost, I need to thank my committee members for their invaluable guidance and support throughout this process: Felix Oberholzer-Gee, Jan Rivkin, Juan Alcacer, and Victor Bennett. Thank you for your persistence and belief in me through the ups and downs of codifying my research interest and generating research ideas. Though I eventually came nearly full circle back to the interesting phenomenon I wanted to study upon my application to graduate school, I certainly took the scenic route to get to here. Each of you made investments in me along the way that has helped form and guide my path.

Thank you to my friends and community – both those present at HBS and scattered elsewhere. To those whom I have gotten to know through HBS and have seen me through this process from up close, our daily and weekly conversations about research and life have guided me through my mid-to-late twenties and into a new decade of life that looks very different from the one I had when I first walked onto this campus in 2010. We have shared struggles and laughs as we have come to appreciate this unique and special place called HBS. And to those whom I have gotten to know in childhood, through Princeton, at Merrill Lynch and in New York, and now in Chicago, I am incredibly grateful for your friendship and support as I pursued a degree and a path that not many people (including myself until just before I applied for it) even knew existed.

Further, your gracious nods as I have attempted to describe what I do with my time and why it takes so long are ever appreciated.

Thank you to my family. Though I suppose I did not really have any choice in my family members, it has been in progressing through adulthood that I have become increasingly appreciative of the ones that get I call my own. Throughout my entire life, you have been constant cheerleaders for each of my pursuits and have believed in me when I needed it most. For that, I am forever grateful.

Last but not least, the biggest thank you to my husband, Brad. I will never be more thankful for a research project that did not go anywhere than for the one that led me to the conference where we met. You have enriched my life in more ways than I can count. You have become my constant confidant and a tireless supporter of and encourager to me as I have progressed in this program and am now marking its completion. My best memories of my time at HBS will undoubtedly be those that I shared with you.

INTRODUCTION

As industries evolve, the practices necessary to obtain and maintain a competitive advantage may also evolve. Therefore, firms' abilities to select and implement these practices can be critical to their success. However, there seems to be wide variation in the internal practices of similar firms. The aim of my dissertation is to understand at a granular level both the fundamental drivers of practice choice and more specifically how practice changes occur within firms. In doing so, I gain a better understanding of why firms differ in their firm practices and why some firms are better than others at navigating transitions and maintaining competitiveness. Specifically, my dissertation is comprised of three large-scale empirical studies that examine specific attributes of and contextual factors within the firm that impact the adoption and implementation of specific management and organizational practices.

In the first essay of my dissertation, my co-authors (Victor Bennett and Raffaella Sadun) and I use the World Management Survey, a large dataset measuring management practices of firms located around the world, to consider how firm ownership impacts the management practices adopted in firms. We find that founder CEO firms have the lowest management scores of all owner-manager type firms and that these low scores may be attributed both to a lack of understanding by founders about the weakness of their practices and to non-pecuniary benefits associated with weak practices accruing to founders.

In the second and third essays of my dissertation, I compile and use a unique dataset from a large North American retail chain that changed the way it restocked its stores to evaluate how and why an identical new practice implementation differs across similar stores.

In the second essay, I consider how a restocking team's prior experience with the old restocking practice impacts its ability to perform and learn the new practice. I find that teams where employees have greater exposure to the old practice perform significantly worse than other teams at the outset which is consistent with the notion that these teams experience 'competency traps'. However, these same teams also learn more quickly, which I suggest may be the result of increased efficiency in communication and coordination within these restocking teams.

In the third essay, I consider how the use of pilots to transfer the new practice impacts performance. I propose and subsequently find evidence that that pilots provide an opportunity for teams to learn from other's experiences in implementing the new practice. This vicarious learning from the pilot store occurs locally based on the established divisions within the firm. Stores learn both from their own experience as well as the pilot store's experiences with the new practice and performance feedback moderates this relationship. The weight stores place on each type of learning – experiential and vicarious – adjusts each week based on relative performance and also depends on the contextual similarity between the pilot and nonpilot store. These findings suggest that the organization of multi-unit firms into divisions and the choice of pilots when implementing new practices have a significant impact on stores' abilities to adapt and change over time.

Combined, the three essays of my dissertation highlight the variation that exists in internal practices of firms and the importance of understanding these variations when examining performance differences among similar firms.

Essay 1 | **Are Founder CEOs Good Managers?**

Victor Manuel Bennett
Megan Lawrence
Raffaella Sadun

ABSTRACT

We investigate the management practices adopted by firms where the founders are also the CEOs using data from the World Management Survey. We find that founder CEO firms have the lowest management scores of any owner-manager pair type and that this difference is associated with significant performance differentials. We propose three possible reasons for the managerial gap of founder CEO firms: a) informational problems preventing a clear understanding of the weakness of their firms' managerial practices; b) institutional factors dampening the incentive to adopt managerial practices; and c) non-pecuniary returns to potentially inefficient but power-preserving practices. The findings presented in the paper provide support for a) and c), while we do not find evidence that the management practices of founder CEO firms vary with respect to the characteristics of the institutional environments in which they are embedded.

I. INTRODUCTION

There is remarkable variation in the practices by which seemingly similar firms are managed (Bloom and Van Reenen 2007). Those differences have been attributed to a wide variety of industry, firm, and managerial characteristics including competitive pressure (Hermalin 1994; Bennett 2013), psychological traits (Galasso and Simcoe 2011; Malmendier and Tate 2005) or personal "style" of the CEO who leads the organization (Bertrand and Schoar 2003), and the ownership structure of the firm (Morck, Shleifer, and Vishny 1988).

In this paper we study the adoption of basic management practices in firms in which the CEO of the firm and its founder are one-and-the-same – which we define as 'founder CEO' firms in what follows. While founder CEOs are typically portrayed as highly extrinsically and intrinsically motivated individuals (Jensen and Meckling 1976; Wasserman 2006), it is unclear whether they should necessarily serve as top managers of their firm. There are several reasons why founders may not be the best top managers. First, the skills needed to create a new venture may not necessarily coincide with capabilities needed to lead the firm through more advanced phases of growth and expansion.[1] Furthermore, founder CEOs might be reluctant to adopt practices that standardize the operations of the firm, since these practices reduce the idiosyncratic and personalized aspects of the entrepreneur's role (Rajan 2012) and the private benefits of control associated with them (Bandiera, Prat, and Sadun 2013).

We investigate these issues using the World Management Survey (WMS), an international dataset providing detailed information on the management practices for a large sample of medium and large manufacturing firms (Bloom et al. 2014; Bloom and Van Reenen 2007) in 32

[1] This viewpoint is supported by the fact that venture capital firms and private equity firms frequently replace founders with professional managers (Hellmann and Puri 2002).

countries. The management processes surveyed in the WMS are akin to managerial "best practices" and have been found to be strongly and causally related to superior firm performance (Bloom and Van Reenen 2007; Bloom et al. 2012).

The WMS includes a large number of founder and non-founder CEOs firms of similar ages and sizes within the same industries and countries. Although we cannot estimate causal effects of being led by a founder CEO, the richness of the data allows us to examine the conditional correlation between management and the founder CEO status of the company while controlling for a large set of potentially confounding covariates suggested by theory and earlier empirical investigations such as firm age, size, average skills of the workforce, country of operation, and main industry of activity.

We start our analysis by reporting three main stylized facts. First, firms led by founder CEOs have lower management scores relative to other forms of concentrated and dispersed ownership. Second, the association between management and firm performance in founder CEO firms is positive and significant, similar to what is generally found for other ownership types. This positive association suggests both that the lower level of management quality in founder CEO firms is likely to result in worse firm performance and that lower management scores among founder CEO firms are not due to the fact that these firms have lower returns to management. Third, firms led by founder CEOs experience significant improvements in their management practices upon a change of ownership, and these improvements are generally much larger than what is found for other ownership transitions.

A natural question arising from these findings is: why are firms led by founder CEOs not adopting performance-enhancing managerial processes, or replacing themselves with managers who do? We present three not-necessarily-mutually-exclusive possible classes of explanations for the persistence of poor management practices at firms with founder CEOs despite the performance penalty: a) that founder CEOs are unaware of their managerial gaps; b) that environmental or institutional variables make it more costly or less attractive for founder CEOs to hire more capable managers to replace themselves, or to select practices consistent with the process of standardization needed to attract external capital (Rajan 2012); and c) that the adoption of formalized managerial processes may interfere with the founders' ability to pursue non-pecuniary benefits of control, such as investing in a pet project or hiring people based on personal or family affiliations. The initial findings presented in the paper provide support for a) and c), but we do not find evidence that founder CEO firms are systematically different according to the quality of the institutional environments in which they are embedded.

Our findings face several limitations. First, the nature of the firms included in the WMS data (companies between 50 and 5000 employees) significantly dampens our ability to analyze the role of founder CEOs on organizations in their early stages of life and/or managers in the early years of their tenure, which may both be more salient to the entrepreneurship literature. Second, the nature of our data does not allow us to estimate the *causal* effect of founder CEOs on management adoption and firm performance; rather we present simple conditional correlations. Relatedly, the lack of information on CEO skills, preferences and experiences does not allow us to look in more detail at the heterogeneity within different types of founder CEOs.

The paper is structured as follows. In Section 2 we provide a description of the WMS data. In Section 3 we explore the differences in management practices between firms led by founder CEOs and firms and all other forms of leader-ownership. In Section 4 we explore the relationship between management and firm performance. In Section 5 we present the analysis of the possible drivers of the managerial differences across ownership types. Section 6 concludes.

II. DATA

II.1SURVEY METHODOLOGY

To measure the presence of basic management practices, we use the World Management Survey (WMS), which was collected using a methodology first described in Bloom and Van Reenen (2007). The survey is based on an interview-based evaluation tool that defines and scores from 1 ("worst practice") to 5 ("best practice") across 18 key management practices. Appendix Table 1 lists the management questions and also gives some sense of how the responses to each question are mapped onto the scoring grid.[2]

The evaluation tool attempts to measure management practices in three key areas. First, *monitoring*: How well do organizations monitor what goes on inside the firm and use this information for continuous improvement? Second, *targets*: Do organizations set the right targets, track the right outcomes, and take appropriate action if the two are inconsistent? Third, *incentives/people management*: Are organizations promoting and rewarding employees based on performance, prioritizing careful hiring, and trying to keep their best employees?[3]

The methodology gives a firm a low score if it fails to track performance, has no effective targets, does not take ability and effort into account when deciding on promotions (e.g. completely tenure-based) and has no system to address persistent employee underperformance.

---

[2] For the full set of questions for each sector (manufacturing, retail, schools and hospitals) see www.worldmanagementsurvey.org.

[3] These practices are similar to those emphasized in earlier work on management practices, by, for example, Osterman (1994), Ichniowski, Shaw, and Prennushi (1997) and Black and Lynch (2001).

In contrast, a high scoring organization frequently monitors and tries to improve its processes, sets comprehensive and stretching targets, promotes high-performing employees and addresses (by re-training/rotating and, if unsuccessful, dismissing) underperforming employees.

The survey design included teams of MBA-type students with business experience conducting the interviews with the plant managers in their native languages. Plant managers were purposely selected, as they were senior enough to have an overview of management practices but not so senior as to be detached from day-to-day operations. The survey is based on a double-blind methodology. First, managers were not told they were being scored or shown the scoring grid. They were told only that they were being "interviewed about their day-to-day management practices." To do this, the interviewers asked open-ended questions,[4] and continued with open questions focusing on specific practices and trying to elicit examples, until the interviewer could make an accurate assessment of the firm's practices.[5] Second, the interviewers were not told anything in advance about the organization's performance; they were provided only with the organization's name, telephone number, and industry.

The dataset includes randomly sampled medium-sized firms (employing between 50 and 5,000 workers) in the manufacturing sector. The sampling frame was drawn in such a way that the firms sampled for each country are representative of the distribution of medium-sized

[4] For example, on the first monitoring dimension in the manufacturing survey, the interviewer starts by asking the open question "Could you please tell me about how you monitor your production process?" rather than a closed question such as "Do you monitor your production daily [yes/no]?"

[5] For example, the second question on that monitoring dimension is "What kinds of measures would you use to track performance?" rather than "Do you track your performance?" and the third is "If I walked around your factory what could I tell about how each person was performing?". The combined responses to the questions within this dimension are scored against a grid that goes from 1, which is defined as "*Measures tracked do not indicate directly if overall business objectives are being met. Tracking is an ad hoc process (certain processes aren't tracked at all)*," to 5, which is defined as "*Performance is continuously tracked and communicated, both formally and informally, to all staff using a range of visual management tools.*"

manufacturing firms across a variety of different databases. The survey achieved a response rate of about 50% through a combination of government endorsements and internal managerial efforts. Reassuringly, responses were uncorrelated with the (independently collected) performance measures for the firm (see Bloom et al. 2014 for details).

The dataset also includes a series of "noise controls" on the interview process itself (such as the time of day and the day of the week), characteristics of the interviewee (such as tenure in firm), and the identity of the interviewer (a full set of dummy variables for the interviewer to account for any interviewer bias). In some specifications we include these variables to control for measurement error. The data was also internally validated through silent monitoring of the interviews (whereby a second person listening in on a phone extension independently scored the interview), and repeat interviews (using a different interviewer and a second plant manager within the same firm). In both cases, the comparisons suggested a high level of consistency across different interviewees and interviewers (see Bloom et al. 2014 for details).

II.2 OWNERSHIP

Firms are classified in several different ownership categories using information collected during the survey and are subsequently cross-checked against public accounts and web searches. This process first determines whether any individual person, group of individuals or organization owns more than 25.01% of the shares of the company. If this is not the case, the firm is classified as owned by "Dispersed Shareholders". If a single group of individuals or organization owns more than 25.01% of the shares of the company, the firm is subsequently classified in the following categories according to the nature of the controlling individuals/organization: "Founder" (the owner coincides with the person who founded the firm); "Family" (the owner/s are affiliated with the family of the firm's founder); "Private Equity"; "Private Individuals"; "Managers"; "Government". The firm is classified in the "Other" category if the ownership type does not match any of the above categories (this typically happens for country specific ownership types, such as foundations in Germany). When a founder or a family owns the firm, we further distinguish between the cases in which the CEO is the founder him/herself or is affiliated with the owning family.

In what follows, we will focus most of the discussion on the difference between firms that are owned and run by a founder CEO, which represent in total 18% of the sample, and all the other types of ownership. Table 1 presents a detailed breakdown of the frequencies of founder CEO firms included in the sample according to their ownership type across the 32 countries included in the sample. Clearly, founder CEO firms are much more likely to be found in developing countries relative to more developed economies—the fraction of founder CEO firms across OECD economies is 11% vs. 30% in non-OECD countries.[6] Therefore, in our analysis we will primarily examine within country comparisons, in order to allay the concern that the differences in management practices across firms may capture unobserved country characteristics.

[6] This fact is not surprising given that many founder CEO successions are associated with growth milestones (Wasserman, 2003), and developing economies have many more small firms (Hsieh and Olken 2014).

TABLE 1

Firm Ownership Across Countries

Sample	All	All other ownership	Founder CEO
Argentina	566	471	95
Australia	470	442	28
Brazil	1,145	754	391
Canada	418	368	50
Chile	543	471	72
China	761	601	160
Colombia	170	114	56
Ethiopia	131	90	41
France	610	571	39
Germany	608	592	16
Ghana	107	54	53
Greece	272	222	50
India	921	529	392
Italy	310	252	58
Japan	172	168	4
Kenya	184	134	50
Mexico	524	424	100
Mozambique	85	59	26
New Zealand	149	135	14
Nicaragua	97	77	20
Nigeria	118	55	63
Poland	364	330	34
Portugal	311	252	59
Republic of Ireland	161	127	34
Singapore	373	308	65
Spain	213	194	19
Sweden	377	369	8
Tanzania	150	102	48
Turkey	332	173	159
United Kingdom	1,332	1,225	107
United States	1,393	1,267	126
Zambia	68	46	22
Total	**13,435**	**10,976**	**2,459**

III. MANGEMENT PRACTICES IN FOUNDER CEO FIRMS

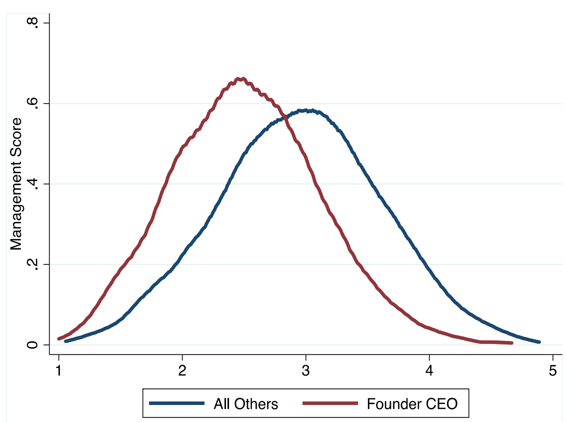
III.1 CROSS-SECTIONAL ANALYSIS

In this section we examine the differences in management practices across different ownership types, focusing, in particular, on firms owned and managed by their founder.

Table 2 shows summary statistics for the overall sample, and the raw comparisons between founder CEO firms and the rest of the ownership categories. The first three rows of Table 2 show that founder CEO firms on average appear to be much less likely to have adopted the basic managerial practices included in the WMS. This gap is significant when we consider the overall management score as well as when we distinguish between the operational questions (*monitoring* and *target setting*) and the *people management* questions asked in the survey.[7] Looking beyond sample means, Figure 1 presents a kernel density plot of management scores for founder CEO firms and firms with other ownership types. The graph shows that the lower average is not due to a tail of firms with low management bringing down the average but rather that the entire mass of the distribution is shifted to the left.

FIGURE 1

Kernel Density Plot of Management Scores for Founder CEO firms and all other Ownership Types



[7] The gap in management scores between founder CEO firms and other ownership types is still evident when we use a more granular ownership classification. Figure 1 in the Appendix plots the raw average management scores across the finer ownership classifications introduced in section 2.B. Founder CEO firms have the lowest average management scores even relative to the second-lowest category, family firms managed by a family CEO. The difference between the two types of ownership is significant at the 1% level, and remains so even when we control for country and industry (3 digit SIC) fixed effects.

TABLE 2

Summary Statistics

Sample	(1) Total	(2) All other ownership	(3) Founder CEO	(4) (2)-(3), p-value
Management	2.873	2.952	2.518	0.434***
	(0.678)	(0.669)	(0.600)	(29.63)
Operations	2.941	3.037	2.515	0.522***
	(0.764)	(0.750)	(0.678)	(31.70)
People	2.736	2.783	2.524	0.259***
	(0.653)	(0.656)	(0.593)	(18.00)
Firm employment	850.382	952.282	395.753	556.5***
	(3821.212)	(4205.027)	(778.492)	(6.53)
Plant employment	270.084	280.909	223.831	57.08***
	(410.197)	(427.679)	(321.067)	(6.09)
Firm age	48.463	52.266	25.297	26.97***
	(42.500)	(44.469)	(11.769)	(20.34)
MNE status	0.404	0.476	0.088	0.388***
	(0.491)	(0.499)	(0.283)	(36.56)
Skills	15.068	15.637	12.644	2.993***
	(16.893)	(17.147)	(15.536)	(7.58)
Observations	13,436	10,977	2,459	

Notes: Table is calculated with simple averages. Column (4) indicates that the differences in raw averages between Founder CEOs and all other ownership are significant at the 1% level across all variables. MNE STATUS is an indicator variable equal to 1 if the firm is a multinational. SKILLS measures the proportion of firm employees (managers and non-managers) with a college degree. MANAGEMENT is the average management score based on responses to the 18 categories assessed in the WMS (Bloom and Van Reeen, 2007). OPERATIONS is the average management score for the set of questions associated with monitoring and target practices. PEOPLE is the average management score for the set of questions associated with HR practices within the firm.

Clearly, management is not the only dimension along which founder CEO firms differ from the other ownership types included in the WMS. Although the criteria for inclusion in the management survey skew the distribution towards larger firms, it is still the case that founder CEO firms are smaller and younger than the other firms in the sample. Founder CEO firms are also less likely to be part of a domestic or foreign multinational and have, on average, a smaller fraction of employees with a college degree. To understand the extent to which the differences in management scores between founder CEO firms vs. other ownership types can be accounted for by these observable firm characteristics—which are typically associated with differences in management practices (e.g. Bloom, Sadun and Van Reenen 2015)—in Table 3 we show the conditional correlation between management and the founder CEO dummy controlling for a progressively larger set of controls (standard errors clustered at the firm level are shown in parentheses under the coefficients). To the extent that these differences are endogenous to ownership, the resulting estimates will provide a lower bound to the causal effect of the founder CEO dummy.

TABLE 3

Founder CEO Management

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sample	All	All	All	All	All	Non-OECD	OECD
Founder CEO	-0.412***	-0.254***	-0.266***	-0.162***	-0.138***	-0.128***	-0.148***
	(0.022)	(0.021)	(0.021)	(0.021)	(0.019)	(0.025)	(0.031)
ln(Firm employment)		0.233***	0.235***	0.194***	0.176***	0.179***	0.175***
		(0.008)	(0.008)	(0.008)	(0.007)	(0.011)	(0.010)
ln(Firm age)			-0.063***	-0.038***	-0.041***	-0.015	-0.043***
			(0.013)	(0.013)	(0.012)	(0.036)	(0.013)
Skills				0.133***	0.120***	0.138***	0.106***
				(0.007)	(0.006)	(0.009)	(0.009)
MNE status				0.364***	0.325***	0.341***	0.319***
				(0.019)	(0.017)	(0.029)	(0.021)
Constant	-0.711***	-1.843***	-1.613***	-1.924***	-3.894***	-3.667***	-2.375***
	(0.097)	(0.112)	(0.121)	(0.119)	(0.623)	(0.195)	(0.606)
Observations	13,436	13,436	13,436	13,436	13,436	4,877	8,559
Adjusted R-Squared	0.182	0.287	0.289	0.337	0.450	0.477	0.367
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm employment	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm age	No	No	Yes	Yes	Yes	Yes	Yes
Skills	No	No	No	Yes	Yes	Yes	Yes
MNE status	No	No	No	Yes	Yes	Yes	Yes
Noise	No	No	No	No	Yes	Yes	Yes

Notes: Dependent variable is the management z-score. All columns estimated by ordinary least squares (OLS) with standard errors clustered at the company level (due to inclusion of a subset of panel firms). Columns (1) - (5) use the entire sample for estimation; Columns (6) and (7) repeat specification (5) for non-OECD and OECD countries separately. Country controls are a full set of country dummies for the countries in which the headquarters of each firm is located (which may be different from the country in which the interviewed plant manager is located for the case of multinational firms). Industry controls are SIC three-digit dummies. Firm employment, firm age, skills and MNE status are included and described in Table 1. Noise controls include the duration of the interview and an indicator for the specific person conducting the interview. *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

The dependent variable in all regressions presented in Table 3 is the firm-level average management score, aggregated across all questions and standardized. Column 1 shows that the relationship between lower management scores and founder CEOs is significant when comparing firms within countries. The difference is large (0.412 of a standard deviation) and significant at the 1% level. Column 2 adds industry (SIC 3 dummies) and log firm employment to control for size and the different distribution of *Founder CEO* firms across sectors. Since larger firms tend to be better managed on average, adding firm size reduces the magnitude of the coefficient on the *Founder CEO* dummy from 0.412 to 0.254, but it remains significant at the 1% level. In Column 3 we add a control for the log of firm age to verify the extent to which the management gap may be driven by firm age, which Table 2 shows to differ significantly across ownership types. Even looking across firms of a similar age, the *Founder CEO* dummy remains of a similar magnitude and significance. In Column 4 we add controls for fraction of employees (managers and non-managers) with college degrees, and multinational status, two variables that are empirically correlated with higher management scores and are systematically less prevalent in founder CEO firms. As a result, the coefficient on the *Founder CEO* dummy is almost halved, becoming 0.162, but the coefficient remains significant at the 1% level. Finally, in Column 5, our baseline specification going forward, we add a set of interview noise controls including interviewer identity and length of the interview. In this specification, the magnitude of the coefficient on the founder CEO dummy lowers to 0.138. Finally, because of evidence that developed countries have higher management practices, on average, in columns 6 and 7 we look at differences across non-OECD and OECD countries and find the results to be remarkably similar, and statistically indistinguishable, across the two subsets.

Overall, the multivariate analysis shows the existence of a managerial gap in founder CEO firms relative to other ownership types which is not fully accounted for by differences in firm country of location, industry of activity, firm size, age, or skills. Using the estimates from Table 3, column 1 and column 5, the analysis reveals that observable firm, industry characteristics, and interview noise are able to account for about 67% of the within country difference between founder CEO firms and other forms of ownership ((0.412-0.138)/0.412), with the rest still being captured by the *Founder CEO* dummy. To further explore the extent to which other unobservable firm characteristics – rather than founder CEO ownership and control - may account for this remaining gap, we turn to analyzing changes in management over time across different types of ownership.

III.2 PANEL ANALYSIS

About 2,844 firms included in the WMS were interviewed more than once over time and, of these, 905 also experienced a change in ownership type. Of these, 167 (of the 487 total founder CEO firms in the subsample of 2,844 firms) classified as founder CEO firms in their first appearance in the WMS dataset transition to a different form of ownership. In this section, we exploit this specific sample with panel management data to further explore the extent to which the managerial gap examined in Section 3.A can be traced back to founder CEO ownership, rather than to other unobservable fixed firm characteristics.

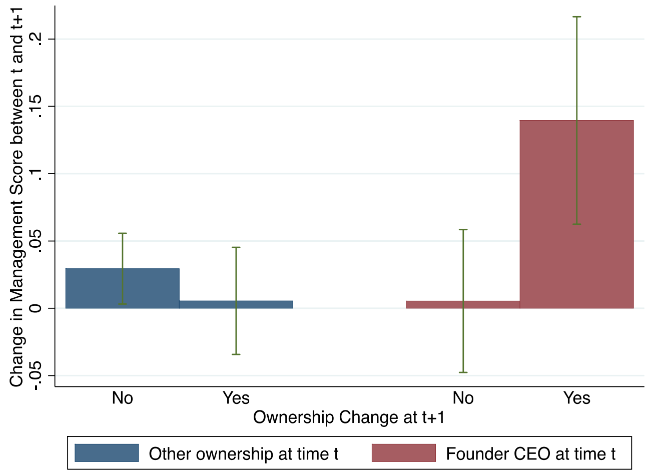
More specifically, we examine whether firms that where initially—i.e. at the time of their first appearance in the WMS data—owned and managed by their founder and experienced a change in ownership before their subsequent appearance in the WMS data saw an improvement in their management scores relative to firms that did not experience an ownership change. Ownership changes are likely to be endogenous—firms are typically acquired on the basis of unobservable characteristics including their productivity or potential for improvement. Therefore, to control for the possibility that the post-acquisition management scores might reflect dynamics unrelated to the change in ownership, we set up this comparison using a difference-in-difference approach, comparing the change in management scores experienced by initial founder CEO firms transitioning to other ownership types (167 firms) to the change in management scores experienced by firms that were initially classified in other ownership categories and also experienced a change in ownership (738 firms).

The identification assumption underlying this comparison is that the unobserved factors leading to an ownership change in founder CEO firms are similar to those leading to an ownership change in other types of firms. To gauge the empirical relevance of this assumption, we investigated the relationship between a dummy capturing the ownership change between two distinct waves of the WMS and a basic set of firm level controls. Reassuringly, the results (presented in Appendix Table 2) show that changes in ownership are not significantly correlated with the initial level of management in both types of transitions, nor with firm size. However, firm age and MNE status both appear to be positively and significantly correlated with changes in ownership for founder CEO firms, but not for the other ownership types. Therefore, while we do not find evidence that founder CEO firms undergoing an ownership change are differentially selected on the basis of their overall management scores relative to other ownership types, we cannot entirely rule out differential selection based on other observable firm characteristics, which may be associated with future changes in management.

With this caveat in mind, we report the graphic result of the difference-in-difference in Figure 2. The bars show the change in management score between two periods, t (the first time a firm appeared in the WMS) and $t+1$ (the last time a firm appeared in the WMS), for four classes of firms. On the left hand side of the graph, we focus on firms that at time t were not owned by a founder CEO and distinguish between those that at $t+1$ had not experienced an ownership change (far left bar in the graph, 1619 firms), and those that had experienced an ownership change (second bar from the left, 738 firms). The left-hand side comparison indicates that there is no significant change in the management scores for firms initially classified in the non-founder CEO category, regardless of ownership changes. On the right hand side of the graph, we repeat the same classification for firms that were at time t classified as founder CEO firms, and distinguish between those that remained classified as such at time $t+1$ (third bar in the graph, 320 firms), and those that instead had transitioned to a different ownership type at time $t+1$ (far right bar in the graph, 167 firms).

FIGURE 2

Changes in Management Score Based on Ownership Changes



The graph shows average change in management score for each of four categories of ownership observed in the WMS panel dataset: non-founder CEO firms with no change in ownership (1619 firms), non-founder CEO firms with a change in ownership (738), founder CEO firms with no change in ownership (320), and founder CEO firms with a change in ownership (167). The error bar values denote 5% confidence intervals for each category.

The graph shows that while the average change in management score between *t* and *t+1* is not distinguishable from zero for founder CEO firms that did not experience a change in ownership, those firms that began with a founder CEO and had transitioned to a different ownership type by *t+1* experienced a significant increase in their management score.

Although the graph is based on raw data, these results are robust to the inclusion of country and industry dummies, firm characteristics and interview noise, as shown in Table 4. Just like in Figure 2, the dependent variable in all columns of Table 4 is the raw change in the average management score between *t* and *t+1*. In column 1 we include as dependent variables only country dummies and an indicator for whether the ownership status changed. The results suggest that change in ownership *per se* is not associated with a significant change in management practices. In Column 2, we add an indicator for whether the ownership type was *Founder CEO* in period *t*, and we find that the coefficient is positive but statistically insignificant, suggesting that founder CEO firms overall did not experience large improvements in management between the two time periods. In Column 3, we include an interaction between the indicators for having a founder CEO in period *t* and a change in ownership prior to time *t+1*. This positive and significant coefficient shows that firms that used to be owned and run by their founder experience large gains in their management score when these firms experience a change in ownership prior to time *t+1*. The magnitude of the coefficient in the interaction is 0.171 which is 28% of the standard deviation in founder CEO score and significant at the 1% level. The magnitude and significance of the coefficient is robust to the inclusion of industry dummies (column 4), and other firm and noise controls (column 5), including the dummy capturing MNE status and firm age.

TABLE 4

Impact of Ownership Changes on Management Scores

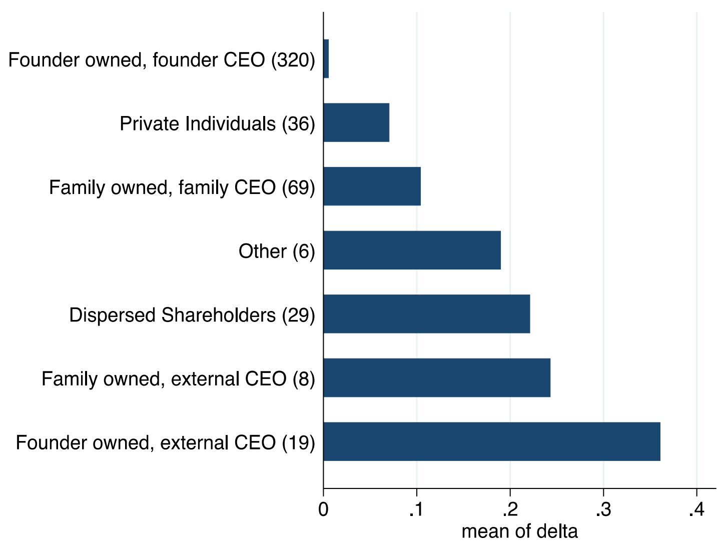
	(1)	(2)	(3)	(4)	(5)
Dependent Variable			Change in Management Score		
Ownership Change	0.016	0.015	-0.017	-0.015	-0.001
	(0.028)	(0.028)	(0.031)	(0.032)	(0.032)
(initial) Founder CEO		0.046	-0.016	-0.015	0.011
		(0.031)	(0.038)	(0.040)	(0.041)
Ownership Change * (initial) Founder CEO			0.171***	0.153**	0.190***
			(0.064)	(0.066)	(0.066)
Constant	-0.060	-0.069	-0.060	-0.083	-0.897***
	(0.047)	(0.048)	(0.048)	(0.051)	(0.225)
Observations	2,844	2,844	2,844	2,844	2,844
Adjusted R-Squared	0.008	0.009	0.010	0.016	0.083
Country dummies	Yes	Yes	Yes	Yes	Yes
Industry dummies	No	No	No	Yes	Yes
Firm employment	No	No	No	No	Yes
Firm age	No	No	No	No	Yes
Skills	No	No	No	No	Yes
MNE status	No	No	No	No	Yes
Noise	No	No	No	No	Yes

Notes: Dependent variable is the change in management score between the first and the last time a firm was interviewed for the WMS. Therefore, only firms who have been administered the survey 2 or more times are included in this estimation. All columns are estimated using OLS and robust standard errors. OWNERSHIP CHANGE is an indicator variable equal to 1 if an ownership change occurred between the first and last time the focal firm was interviewed for the WMS. (INITIAL) FOUNDER CEO is equal to 1 if the firm had Founder CEO ownership the first time the WMS was administered and equal to 0 otherwise. *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

Overall, these results suggest that the differences in management scores discussed in Section 2 are tightly related to the identity of the CEO, rather than being driven by unobserved characteristics of the firms led by founder CEOs. To further illustrate this point, in Figure 3 we break down the changes observed in founder CEO firms at time *t+1* according to the detailed type of ownership at time *t+1*. The average change in management scores is positive across all transitions. Interestingly, the largest change appears when the founder remains the main owner of the firm but an external manager takes the top position. This suggests that it is the presence of the founder in an active operational role in the company that potentially dampens management adoption, rather than founder ownership *per se.*

FIGURE 3

Changes in Management Score for Firms Originating with Founder CEOs



The graph shows the change in management score for firms that were surveyed more than once in the WMS data and were owned and managed by Founder CEOs in the first survey wave in which they appeared. The bars display the average change in management score for each type of ownership transition, indicated in the last observation in the WMS data (as well as the changes in management score for those Founder CEO firms that experienced no transition - the first row). The number of observations of each type of transition (as well as the non-transition group) is shown in parentheses next to the ownership type.

IV. DOES MANAGEMENT MATTER IN FOUNDER CEO FIRMS?

A growing body of research has documented the presence of large and significant performance implications for the managerial practices investigated in the WMS (Bloom et al. 2014; 2013; Bloom and Van Reenen 2007). However, one possible explanation behind the managerial gap explored in Section 3 is that formalized managerial processes might be relatively less important for the performance of founder CEO firms. For example, founder CEOs might be able to substitute for formalized practices with other unobservable managerial skills, such as their charisma, connections, or intrinsic motivation.

We investigate this issue in Table 5, where we estimate a simple production function—log sales as a function of the total number of employees, capital and materials, all drawn from published accounts drawn from the accounting database ORBIS using the following specification:

$$y_{itsc} = \alpha FounderCEO_{it} + \beta Management_{it} + \gamma FounderCEO_{it} * Management_{it} + F_{it}\theta + \delta \mathrm{e}_{it} + \vartheta m_{it} + \mu k_{it} + \zeta_s + \tau_t + \rho_c + \varepsilon_{itsc}$$

where y, e, m, k represent the natural logarithm of, respectively, firm level sales, employment, materials and capital; F the set of firm level controls employed in earlier tables; and $\zeta_{s,}\ \tau_t$ *and* ρ_c denote industry, time and country fixed effects. Since we use repeated cross sections for each firm, errors are clustered at the firm level across all columns. The key parameter in this specification is γ, which allows us to evaluate whether the relationship between management and performance is systematically different for founder CEO firms relative to other ownership types.

TABLE 5

Performance of Founder CEO firms

	(1)	(2)	(3)	(4)	(5)	(7)
Dependent Variable	ln(sales)	ln(sales)	ln(sales)	Change in ln(sales)	ROCE	ROA
Founder CEO	-0.094**	-0.082*	-0.079*	-0.000	-0.174	14.310
	(0.046)	(0.045)	(0.044)	(0.009)	(0.979)	(59.670)
Management		0.093***	0.092***	0.006**	1.035***	52.259**
		(0.015)	(0.015)	(0.002)	(0.356)	(22.746)
Founder CEO*Management			0.006	0.001	-0.658	-68.110
			(0.048)	(0.007)	(0.728)	(43.172)
ln(firm employment)	0.628***	0.616***	0.616***		1.500***	65.538**
	(0.023)	(0.023)	(0.023)		(0.440)	(27.632)
ln(materials)	0.226***	0.224***	0.224***		1.149***	52.683**
	(0.014)	(0.014)	(0.014)		(0.374)	(24.274)
ln(capital)	0.246***	0.240***	0.240***		-1.008***	3.456
	(0.016)	(0.015)	(0.015)		(0.336)	(20.129)
Change in ln(firm employment)				0.417***		
				(0.026)		
Change in ln(materials)				0.518***		
				(0.019)		
Change in ln(capital)				0.151***		
				(0.013)		
Constant	2.902***	3.078***	3.076***	-0.068	0.173	-1956.993
	(0.295)	(0.290)	(0.290)	(0.084)	(10.420)	(1684.996)
Observations	9,203	9,203	9,203	8,902	7,677	8,720
Adjusted R-Squared	0.807	0.810	0.810	0.388	0.100	0.089
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes
Firm employment	Yes	Yes	Yes	Yes	Yes	Yes
Firm age	Yes	Yes	Yes	Yes	Yes	Yes
Skills	Yes	Yes	Yes	Yes	Yes	Yes
MNE status	Yes	Yes	Yes	Yes	Yes	Yes
Noise	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The sample used for this table includes only those firms for which sales, employment, capital, ROCE and ROA data could be found in ORBIS and other databases. All columns are estimated using OLS and standard errors clustered at the firm level. *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

Column 1 shows that founder CEO firms tend, on average, to be 9.4% less productive than other ownership types (the coefficient is significant at the 5% level). Column 2 adds to the specification the average management score which, consistent with earlier research, appears to be positive and strongly correlated with productivity (coefficient 0.093, standard error 0.015). This result also shows that, although differences in management are able to account for about 13% of this difference (0.094-0.082/0.094), the *Founder CEO* dummy remains statistically significant at the 10% level. In column 3 we introduce the *Founder CEO*Management* interaction to test for differential slopes by ownership types. We find the interaction to be small and positive, though statistically insignificant at conventional levels. This basic finding is confirmed in columns 4, 5 and 6, where we look, respectively, at one year log changes in sales, ROCE, and ROA as alternative outcome variables.

Overall, we find no support for the hypothesis that management might be a less critical factor in firms led by their founders relative to other ownership types.

V. WHY DO FOUNDER CEOs HAVE LOW MANAGEMENT SCORES?

The persistence of founder CEOs using weaker management practices in light of the positive performance associated with management is a puzzle. If founder CEOs have a stake in the financial performance of the organization, it seems like they would be better served by either adopting performance-enhancing practices, or by replacing themselves with professional managers.

In this section, we explore some of the reasons why we might observe this non-adoption of management practices among founders. First, we investigate whether the managerial gap explored in Section 3 might be due to *informational* constraints, i.e. founder CEOs might simply not know or not be able to recognize the added value of the practices we investigate. Second, founder firms may arise in situations where the incentive to adopt these practices and standardize the business practices of the organization might be dampened by the institutional constraints in which the firms are embedded (Rajan 2012). Third, founders might resist the adoption of formalized management practices because they derive non-monetary benefits of control (Hamilton 2000; Moskowitz and Vissing-Jørgensen 2002) and perceive these processes as a potential obstacle to the pursuit of possible private benefits. We explore these non-mutually-exclusive arguments below.

V.1 INFORMATIONAL CONSTRAINTS

One potential explanation for the wide heterogeneity in adoption of performance-enhancing management practices across firms might be due to problems of *perception*—i.e. founders may underestimate the practices' effect on productivity or overestimate the degree to which they are being implemented in practice (Gibbons and Henderson 2012).

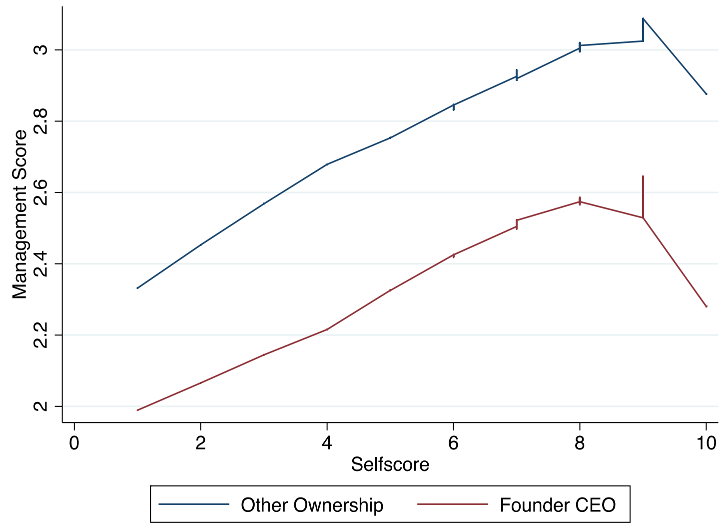
To investigate whether the perception problem might be a possible explanation of the managerial gap documented across founder CEO firms, we exploit a self-reported measure collected at the end of the WMS survey in which managers assess the quality of their own practices on a scale from 1 to 10.[8] Figure 4 plots the average standardized WMS scores associated with the manager self-assessed scores (generated using a non-parametric *lowess* estimator overlaid onto the scatter plot of values) for both founder CEO firms and the other ownership types. The self-assessed own-firm management score and the one obtained through the WMS interviews are positively correlated for all but the highest level of self-assessment, where true score trends down slightly in both cases. Interestingly, however, managers at founder CEO firms tend to systematically overestimate how well managed their firm is—the same level of self-score maps into a systematically lower level of actual management score for founder CEO firms.[9]

[8] The exact wording of the question is: "Ignoring yourself, how well managed do you think the rest of the company is on scale: 1 to 10, where 1 is worst practice, 10 is best practice and 5 is average?".

[9] Because the phrasing of the question rules out the manager evaluating his/herself, these results do not seem to be consistent with personal overconfidence. The results may be consistent with hiring policies resulting in less experienced managers or with weak performance monitoring policies that result in managers having a weak idea of what works, however.

FIGURE 4

Manager Self-Score of Firm Management Compared with WMS Management Score



The graph shows the result of a lowess estimator of self-responses of the interviewed plant manager when asked to indicate his/her impression of firm management (on a scale of 1-10) as compared to the management score derived from the WMS interview.

To look in more detail at the relationship between actual and self-assessed scores across ownership types, we define an "awareness" metric in the following way. First, we categorize each firm according to its quintile in the actual management score distribution within its country. Second, we do the same for the self-assessed management quality by country. Third, we define a variable taking values as follows: -1 if the difference between the actual and self-assessed quintile is less than -1, indicating that the manager systematically underestimated the relative quality of his or her firm's management quality; 0 if the difference in the quintiles is between -1 and 1 (included), if the self-assessment was relatively accurate; and 1 if the if the difference between the actual and self-assessed quintile is greater than 1, indicating the manager systematically overestimated the relative quality of his or her firm's own management quality. Table 6 summarizes the values of this variable across different ownership types. Overall, about 57% of the managers appear to have a relatively good idea of where their firm stands in terms of management. About 30% seem to underestimate their firm's relative standing, while 13% overestimate their firm's management quality relative to the actual scores. The distribution of the scores across these three categories of managers, however, is systematically different across ownership types. More specifically, founder CEO firms tend to have a larger fraction of firms that overestimate (22% vs. 11%) or have a realistic assessment (64% vs. 55%) of their scores and a much smaller fraction that underestimate their scores (14% vs. 34%).

TABLE 6

Own-Firm Management Self-Assessment by Ownership Type

	Total	All other Ownership	Founder CEO
Underconfidence	3,775	3,448	327
	30.16%	33.90%	13.93%
Realism	7,112	5,608	1,504
	56.81%	55.14%	64.08%
Overconfidence	1,631	1,115	516
	13.03%	10.96%	21.99%
Total	12,518	10,171	2,347
	100%	100%	100%

Notes: Table includes raw number of firms for which underconfidence, realism, and overconfidence were detected in the interviewed plant manager's self assessment of his/her firm's management. To collect the self-score, managers were asked on a scale of 1-10 how they perceived their firms' management proficiency. This data was subsequently divided into quintiles as were the WMS management scores, separately. UNDERCONFIDENCE is classified as having a self-assessment quintile value at least 2 lower than the actual management score of the firm. REALISM is assigned to a firm if the interviewed manager's self-score of the firm's management is within 1 quintile (above or below) the actual management score for the firm. Lastly, OVERCONFIDENCE is a result of a managerial self-score of at least 2 quintiles higher than the firm's WMS management score. Along with the raw number of firms, the percentage of the total firms is included for all firms and, separately, Founder CEO firms and firms under all other forms of ownership.

To see whether these differences in awareness might be able to account for the differences in scores documented in Section 3, we include the "awareness metric" in the specification calculated in Table 3, column 5, and test whether the inclusion of this metric has any sizeable effect on the coefficient measuring the *Founder CEO* dummy effect. The results of this exercise are shown in Table 7. We start with a baseline specification in column 2 where we simply show that the coefficient on the *Founder CEO* dummy is still negative and significant and of similar size in the sample of firms for which the self-assessment metric is available (column 2 compared to column 1).[10] In column 3 we add the awareness metric – which reduces the coefficient on the *Founder CEO* dummy by about 25% (from 0.125 to 0.093), but the coefficient is still sizeable and significant at the 1% level. In columns 4 to 7 we repeat the same experiment for firms in non-OECD (columns 4 and 5) and OECD countries (columns 6 and 7). In both cases, the coefficient on the *Founder CEO* dummy remains negative and significant; however, the reduction in its coefficient when the awareness variable in included is much larger in OECD countries (46% vs. 15%).

[10] The smaller sample is due to the fact that the self-assessment question was introduced in the 2006 WMS wave whereas the whole sample started being collected in 2004.

TABLE 7

Accounting for Awareness of Management Quality on Management

Sample	(1) All	(2) All	(3) All	(4) Non-OECD	(5) Non-OECD	(6) OECD	(7) OECD
Founder CEO	-0.138***	-0.125***	-0.093***	-0.125***	-0.106***	-0.122***	-0.068**
	(0.019)	(0.019)	(0.017)	(0.025)	(0.023)	(0.030)	(0.026)
ln(Firm employment)	0.176***	0.172***	0.137***	0.180***	0.148***	0.169***	0.132***
	(0.007)	(0.007)	(0.007)	(0.011)	(0.010)	(0.010)	(0.009)
ln(Firm age)	-0.041***	-0.033***	-0.024**	-0.013	-0.015	-0.034**	-0.023**
	(0.012)	(0.012)	(0.011)	(0.037)	(0.032)	(0.013)	(0.012)
ln(Skills)	0.120***	0.123***	0.095***	0.139***	0.114***	0.111***	0.083***
	(0.006)	(0.006)	(0.006)	(0.009)	(0.009)	(0.009)	(0.008)
MNE status	0.325***	0.337***	0.258***	0.342***	0.271***	0.336***	0.259***
	(0.017)	(0.017)	(0.015)	(0.029)	(0.025)	(0.022)	(0.018)
Awareness			-0.650***		-0.563***		-0.698***
			(0.011)		(0.017)		(0.014)
Constant	-3.894***	-4.110***	-3.352***	-3.694***	-2.955***	-2.827***	-2.119***
	(0.623)	(0.712)	(0.540)	(0.196)	(0.180)	(0.623)	(0.445)
Observations	13,436	12,518	12,518	4,827	4,827	7,691	7,691
Adjusted R-Squared	0.450	0.467	0.592	0.478	0.579	0.392	0.549
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm employment	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm age	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Skills	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MNE status	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Noise	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Dependent variable is the management z-score index. All columns estimated by ordinary least squares (OLS) with standard errors clustered at the company level (due to inclusion of a subset of panel firms). Columns (1) - (3) use the entire data set whereas Columns (4) - (7) test the effect of managerial awareness in non-OECD and OECD countries respectively. *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

Overall, these results suggest that the lower managerial scores of founder CEO firms are associated with managers' systematic lack of awareness of the weakness of their firms' management quality (especially in OECD countries), but this lack of self-awareness does not fully explain the management gap that we find for founder CEOs relative to other ownership types.

V.2 INSTITUTIONS

In this section we explore whether inefficient institutions may be a possible driver of the lower managerial scores of founder CEO firms. The potential role of institutions in shaping the incentive to adopt formalized managerial practices can best be seen in terms of the framework proposed by Rajan (2012) to investigate when and to what extent founders will have the incentive to "standardize" their business practices, i.e. to establish processes that "reduce the idiosyncratic and personalized aspects of the entrepreneur's role". This set-up is useful since the processes considered by Rajan encompass several of the managerial practices included in the WMS, for example: a) formalizing implicit agreements with employees; b) spreading the allocation of responsibilities across functions so that they can be more easily managed by outsiders; and c) introducing strategic planning and information systems so that the information that a CEO needs to make decisions is more easily available.

One of the key insights of Rajan's framework is that the standardization decision creates a fundamental tension for the founder. On one hand, standardization might be necessary to attract external capital. Potential backers may see these practices as tools through which the human capital in the firm, particularly the CEO, becomes more replaceable, reducing risk by making the firm more amenable to external control. On the other hand, the founder might resist standardization precisely because it makes his or her personal human capital less critical and more easily substituted by an external CEO. In this set up, the founder is encouraged to adopt these "standardized" practices to gain access to capital markets. If capital markets are not well developed, the rewards associated with standardization will be reduced for the founder, hence reducing the incentive to incur the loss of personal rents associated with it. For this reason, institutions that support liquid capital markets may, by extension, support the adoption of superior management practices in founder-owned ventures.

Institutions might also have an impact on the standardization decisions even in absence of the need to raise capital through the market. For example, delegation to other talented managers able to guide the firm through the standardization process might be prohibitively costly in countries with poor contractual enforcement (Bandiera et al. 2014). These costs might be based on objective constraints – i.e. heightened risk of expropriation – or subjective perceptions of the associated risks – i.e. lack of trust (Bloom, Sadun, and Van Reenen 2012). Therefore, institutions that lower the costs of contractual enforcement or foster generalized trust may lower the costs of adopting superior management practices.

To investigate these issues we estimate the following model:

$$Management_{itsc} = \alpha FounderCEO_{it} + \beta FounderCEO_{it} * S_c + F_{it}\theta + \zeta_s + \tau_t + \rho_c + \varepsilon_{itsc}$$

Our coefficient of interest is β, which captures the differential effect of different country-specific institutional variables (measured in the country in which the firms' central headquarters (CHQ) are located)[11] for founder CEO firms. If institutions play any role in shaping the adoption of formalized management practices, we would expect β>0, meaning that the gap between founder CEO firms and other forms of ownership would be smaller in more efficient institutional environments.[12]

We also investigate differences across different types of management practices covered by the WMS, by estimating this regression for the overall management score, and separately for the operations (all questions referring to *monitoring* and *target* practices) and *people* (all the questions pertaining to HR management practices) sections of the survey. We are specifically

[11] Headquarters is the level at which the institutional constraints are more likely to influence the decision to adopt management practices (see Bloom et al. 2012 for a similar application). An alternative approach would be to match the plant with the institutional variable measured in the country in which the plants are located. The results shown in this section are virtually unchanged when we use this alternative approach.

[12] Note that all regressions include country dummies. Therefore, we do not estimate the linear correlation between country level institutions and management, but their differential correlation across founder CEO firms and other ownership types.

interested in practices related to managing people as they may most directly shape the founder's ability to retain control over the company. For example, introducing more formalized HR may limit the founder's ability to promote family and friends to positions of power and, more generally, to use promotions to reward personal loyalty (Bandiera et al. 2014).

The results of this analysis are shown in Table 8 (we cluster the standard errors at the CHQ country level throughout). We start in columns 1-3 by using as a rough measure of institutional quality, the log of GDP per capita (PPP adjusted and expressed in constant 2005 USD). The interaction *Founder CEO*ln(GDP per capita)* is not significant across any of the columns. We obtain similarly insignificant results by following Rajan and Zingales (1998) in using a variable capturing differences in standards of financial disclosures by country as a proxy for the founder's ability to attract external capital, which is necessary to providing the incentive to standardize. Similarly, the interaction between the founder CEO dummy and a variable capturing the overall quality of the Rule of Law (Kaufmann, Kraay, and Mastruzzi 2011) in columns 7-9 and a measure of generalized trust developed from the World Values Survey (World Values Survey Association, 2008) in columns 10-12 are also all statistically insignificant.

TABLE 8

Impact of Institutional Context on Founder CEO Management

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent Variable	Management	Operations	People	Management	Operations	People	Management	Operations	People	Management	Operations	People
Founder CEO	-0.131***	-0.120***	-0.020	-0.156***	-0.141***	-0.030**	-0.130***	-0.118***	-0.021	-0.129***	-0.121***	-0.013
	(0.021)	(0.016)	(0.014)	(0.019)	(0.016)	(0.012)	(0.021)	(0.017)	(0.013)	(0.021)	(0.016)	(0.014)
Founder CEO*ln(GDP per Capita)	0.006	0.002	0.002									
	(0.009)	(0.009)	(0.005)									
Founder CEO*(Accounting Standards)				-0.000	-0.001	0.001						
				(0.002)	(0.002)	(0.001)						
Founder CEO*(Rule of Law)							0.000	0.000	0.000			
							(0.001)	(0.001)	(0.000)			
Founder CEO*Trust										0.118	0.008	0.186
										(0.189)	(0.146)	(0.117)
Constant	-3.735***	-2.785***	-1.796***	-3.483***	-2.724***	-1.376***	-3.746***	-2.795***	-1.797***	-3.736***	-2.786***	-1.796***
	(0.347)	(0.261)	(0.205)	(0.382)	(0.304)	(0.231)	(0.342)	(0.259)	(0.203)	(0.347)	(0.261)	(0.204)
Observations	12,386	12,386	12,386	10,888	10,888	10,888	12,386	12,386	12,386	12,386	12,386	12,386
Adjusted R-Squared	0.451	0.438	0.332	0.436	0.420	0.328	0.451	0.438	0.332	0.451	0.438	0.333
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm employment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Skills	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MNE status	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Noise	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All columns estimated by ordinary least squares (OLS) with standard errors clustered at the level of the country in which the firm's CHQ is located. Each interaction variable is tested in 3 columns with 3 different standardized dependent variables: overall management score, operations management score and people management score. GDP per Capita is drawn from the World Bank Develpment indicators, measured in the country in which the firm headquarters is located. Similarly, ACCOUNTING STANDARDS is used as a proxy for financial development in the country where the firm headquarters is located (Rajan and Zingales, 1998). RULE OF LAW is drawn from the World Bank's Doing Business Survey. TRUST is derived from the World Values Survey, and debotes the % of people answerinf "yes" to the question ""Generally speaking, would you say that most people can be trusted or that you can't be too careful?" *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

In conclusion, we fail to find evidence that development, or more specifically the quality of the institutional environment in which firms operate, has a role in explaining the relative gap in management practices of founder CEO firms. This finding holds for the overall management score as well as the score relating to people management practices.

V.3 PRIVATE BENEFITS OF CONTROL

As mentioned above, a possible reason for the lack of adoption of formalized management practices across founder CEO firms is that standardization may directly dissipate the personal rents that the founder enjoys by being at the helm of his or her organization. For example, Hurst and Pugsley (2011) found that over 50 percent of new business owners reported non-pecuniary benefits as a reason for starting their businesses, citing reasons like "wanting flexibility over schedule" and "to be one's own boss" as of first order importance for their choices.[13]

Unfortunately, we do not have information on the different *individual* preferences of the managers included in the WMS sample. Our approach is to instead investigate whether the adoption of management practices varies according to differences in *societal* preferences. A primary candidate for this type of exercise is the strength of family values in the country where the firm's Central Headquarters are located. Using an index derived from several questions included in the World Values Survey,[14] Bertrand and Schoar (2006) show that the strength of family values is highly correlated with the fraction of family firms—including founder CEO firms—in the economy and in general with the organizational structure of firms. In our setting, we hypothesize that strong family values may create an incentive for founder and family CEOs to select and reward employees on the basis of family affiliations rather than through potentially more objective merit-based HR processes, whose adoption is measured in our management index.

[13] That is consistent with Bennett and Chatterji (2015)'s finding that 58 percent of people who considered starting a business did so because they wanted to "be [their] own boss, turn a hobby into a job, or control [their] own schedule".

[14] Bertrand and Schoar (2006) used principal component analysis to combine the answers to five family related questions into a single index. The questions include (1) general importance family in life, (2) parental respect by children, (3) parental duty to their children, (4) importance of obedience as a quality in children and (5) importance of independence as a quality in children. We use the same index as a proxy for family values.

We investigate this idea in Table 9, by including in our baseline regression an interaction between the Family Values Index and the *Founder CEO* dummy. The interaction between the strength of family values and the *Founder CEO* dummy is negative, as expected, but statistically insignificant when we look at the overall management score (column 1). Interestingly, however, the insignificance is entirely driven by the operations questions of the survey. When we focus the index on the *people* section of the survey—i.e. the type of practices that are likely to have a more direct effect on the ability to employ family members as employees—in column 3, we find that stronger family values are associated with significantly lower management scores for founder CEO firms.

TABLE 9

People Management in Founder CEO Firms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
						Low external	High external
			All			capital	capital
Sample						dependence	dependence
	Management	Operations	People	People	People	People	People
Founder CEO	-0.131***	-0.123***	-0.012	-0.011	-0.038***	-0.024	-0.004
	(0.022)	(0.018)	(0.011)	(0.013)	(0.013)	(0.017)	(0.023)
Founder CEO*Family Values	-0.017	0.012	-0.053**	-0.044*	-0.067***	-0.089***	-0.040
	(0.047)	(0.037)	(0.021)	(0.025)	(0.022)	(0.021)	(0.041)
Family CEO					-0.086***		
					(0.015)		
Family CEO*Family Values					-0.031		
					(0.019)		
Founder CEO*ln(GDP per Capita)				0.001			
				(0.005)			
Founder CEO*Trust				0.049			
				(0.126)			
Constant	-3.734***	-2.787***	-1.790***	-1.791***	-1.746***	-0.817***	-1.357***
	(0.348)	(0.261)	(0.206)	(0.205)	(0.201)	(0.262)	(0.270)
Observations	12,386	12,386	12,386	12,386	12,386	5,862	5,006
Adjusted R-Squared	0.451	0.438	0.333	0.332	0.334	0.292	0.341
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm employment	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm age	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Skills	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MNE status	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Noise	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable in columns 1 and 2 are, respectively, the overall management z-score and the operations z-score. The dependent variable in columns 3-7 is the people management z-score. All columns estimated by ordinary least squares (OLS) with standard errors clustered at the level of the country in which the firm's CHQ is located. Coluns 6 and 7 split the sample according the the Rajan and Zigales financial dependence variable (below and above the sample median). Family values is derived from the World Values Survey as described in Bertrand and Schoar (2007) and measured in the country in which the firm headquarters is located. GDP per Capita is drawn from the World Bank Develpment indicators, measured in the country in which the firm headquarters is located. RULE OF LAW is drawn from the World Bank's Doing Business Survey. TRUST is derived from the World Values Survey, and debotes the % of people answerinf "yes" to the question ""Generally speaking, would you say that most people can be trusted or that you can't be too careful?" *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

In the subsequent columns of Table 9 we investigate this result further by looking at its sensitivity with respect to the inclusion of additional country controls and examining various subsamples of the data. In column 4 we simply repeat the specification adding as controls other relevant country characteristics (log GDP per Capita and Trust) and their interaction with the *Founder CEO* dummy, to check whether the proxy for family values might capture other salient country characteristics. The coefficient on *Founder CEO*Family Values* is reduced by about 30%, but it remains large and statistically significant at the 10% level.

Because a great deal of research has investigated the impacts of family CEOs (e.g., Villalonga and Amit 2006) and in fact often conflate founder CEOs with family CEOs (Wasserman 2003), in column 5 we add to the specification an interaction between a dummy denoting *Family CEOs* (i.e. CEOs that are affiliated to the founding family, but belong to later generations relative to the founder) and its interaction with *Family Values*. While the management scores of family CEO firms also appear to be lower in countries with strong family values, differently from founder CEO firms, the interaction is not statistically significant.

In line with Rajan (2012), we explore whether the relevance of family values varies according to the nature of the industry in which the firm operates. In particular, we would expect family values to play a relatively smaller role in industries with high external financial dependence (defined as in Rajan and Zingales 1998). It is in these industries where the need to raise external capital is likely to dominate the personal returns to private control. In line with this hypothesis, in columns 6 and 7 we show that *Founder CEO*Family Values* interaction is significant only in industries with low external financial dependence.[15]

Overall, these results provide suggestive evidence that different considerations besides pure profit maximization—e.g. the value provided by foregoing objective HR processes to hire a family member or a friend in the firm—may play a role in explaining the relatively low adoption of management practices across founder CEO firms, especially with respect to processes aimed at formalizing HR processes for employee selection, reward, and retention.

VI. CONCLUSION

We find evidence that firms led by founder CEOs are significantly less likely to implement basic management practices, even if these practices are associated with better firm performance. We explore the reasons for the differential adoption. Specifically, we investigate three potential causes: a) that founders don't perceive their firms to have a management gap; b) that the institutional environment dampens the incentive to implement superior practices; and c) that non-pecuniary benefits from control counterbalance the lost rents from those worse practices. We find support for both a) and c).

The results shown in this paper are broadly consistent with an emerging literature emphasizing the heterogeneity in growth and motivation of entrepreneurial firms (Hurst and Pugsley 2011; Mullins and Schoar 2013; Bennett and Chatterji 2015) and with managerial studies focusing on the positive association between structured management practices and performance across start-ups (Davila, Foster and Jia 2010). We extend this literature by providing additional evidence of

[15] We also investigated whether the presence of strong family values could affect the returns to management practices by repeating the performance regressions from Table 5 including an interaction term *Management*Family Values*. We find no evidence of a lower return associated with management practices in countries where family values are higher (see Table 3 in the Appendix).

the managerial practices adopted by founder CEO firms, and their relationship with country-specific cultural norms, such as family values, across a wide range of countries and industries.

This paper contributes to the existing literature on the performance of founder CEO firms. In contrast to our paper, several studies report a positive effect of founder CEOs on firm performance (Adams, Almeida, and Ferreira 2009; Fahlenbrach 2009). One possible reason for this discrepancy results from the type of firms used in the analysis. While this paper includes a wide range of private and public firms across several countries, the positive effect of founder CEOs effect is typically derived from the analysis of samples of public US enterprises which may have implemented standardized management practices in order to be able to raise external capital (Rajan 2012) or, more generally, be positively selected relative to representative founder CEO firms.

The persistent managerial gap of founder CEO firms described in this paper suggests that government sponsored programs aimed at fostering entrepreneurial activity may face significant challenges in delivering growth. In particular, our results suggest that – in order to be effective – financial support provided to new enterprises may need to be coupled with effective policies aimed at improving the managerial capabilities of founders and a better understanding of their motivations.

Unfortunately, a paucity of data on key differences in CEO skills, experience, preferences and ability prevent us from exploring in further detail the mechanisms through which founder CEO status affects management practice adoption. We see this as a promising area for further research.

BIBLIOGRAPHY

Adams R, Almeida H, Ferreira D. 2009. Understanding the relationship between founder–CEOs and firm performance. *Journal of Empirical Finance* 16(1): 136–150.

Bandiera O, Prat A, Sadun R. 2013. Managing the Family Firm: Evidence from CEOs at Work. *NBER Working Paper*. National Bureau of Economic Research: Cambridge, MA.

Bennett VM. 2013. Organization and Bargaining: Sales Process Design at Auto Dealerships. Management Science, 59(9):2003-2018

Bennett VM, Chatterji AK. 2015. Why Start-ups Don't Start. *Duke University Working Paper*. Duke University: Durham, NC.

Bertrand M, Schoar A. 2003. Managing with Style: The Effect of Managers on Firm Policies. *Quarterly Journal of Economics*. Oxford University Press 118(4): 1169–1208.

Bertrand M, Schoar A. 2006. The Role of Family in Family Firms. *Journal of Economic Perspectives*. 20(2): 73–96.

Black SE, Lynch LM. 2001. How To Compete: The Impact Of Workplace Practices And Information Technology On Productivity. *Review of Economics and Statistics*. MIT Press 83(3): 434–445.

Bloom N, Eifert B, Mahajan A, McKenzie D, Roberts J. 2013. Does Management Matter? Evidence from India. *Quarterly Journal of Economics*. Oxford University Press 128(1): 1–51.

Bloom N, Lemos R, Sadun R, Scur D, Van Reenen J. 2014. The New Empirical Economics of Management. *Journal of the European Economic Association* 12(4): 835–876.

Bloom N, Sadun R, Van Reenen J. 2012. The Organization of Firms Across Countries. *Quarterly Journal of Economics*. 127(4): 1663–1705.

Bloom N, Van Reenen J. 2007. Measuring and explaining management practices across firms and countries. *Quarterly Journal of Economics* 122(4): 1351–1408.

Dávila A, Foster G, Jia N. 2010. Building Sustainable High Growth Startup Companies: Management Systems as Accelerators. California Management Review. Volume 52.

Fahlenbrach R. 2009. Founder-CEOs, Investment Decisions, and Stock Market Performance. *Journal of Financial and Quantitative Analysis*. Cambridge University Press 44(02): 439–466.

Galasso A, Simcoe TS. 2011. CEO overconfidence and innovation. *Management Science* 57(8): 1469–1484.

Gibbons R, Henderson RM. 2012. Relational Contracts and Organizational Capabilities 23(5): 1350–1364.

Hamilton BH. 2000. Does Entrepreneurship Pay? An Empirical Analysis of the Returns of Self-Employment. *Journal of Political Economy* 108(3): 604.

Hellmann T, Puri M. 2002. Venture Capital and the Professionalization of Start-Up Firms: Empirical Evidence. Blackwell Publishers, Inc. 57(1): 169–197.

Hermalin BE. 1994. Heterogeneity in Organizational Form: Why Otherwise Identical Firms Choose Different Incentives for Their Managers. *RAND Journal of Economics* 25(4): 518.

Hsieh C-T, Olken BA. 2014. The Missing 'Missing Middle'. *Journal of Economic Perspectives* 28(3): 89–108.

Hurst E, Pugsley BW. 2011. What Do Small Businesses Do? *Brookings Papers on Economic Activity*. What do small businesses do? Forthcoming.

Ichniowski C, Shaw K, Prennushi G. 1997. The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines. *American Economic Review*. 87(3): 291–313.

Jensen MC, Meckling WH. 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3(4): 305–360.

Kaufmann D, Kraay A, Mastruzzi M. 2011. The Worldwide Governance Indicators: Methodology and Analytical Issues. *Hague Journal on the Rule of Law*. Cambridge University Press 3(02): 220–246.

Malmendier U, Tate G. 2005. CEO overconfidence and corporate investment 60(6): 2661–2700.

Morck R, Shleifer A, Vishny RW. 1988. Management ownership and market valuation: An empirical analysis. *Journal of Financial Economics* 20: 293–315.

Moskowitz TJ, Vissing-Jørgensen A. 2002. The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle? *American Economic Review*. 92(4): 745–778.

Mullins W, Schoar A. 2013. How do CEOs see their role? Management philosophy and styles in family and non-family firms *NBER Working paper*. National Bureau of Economic Research: Cambridge, MA.

Osterman P. 1994. How common is workplace transformation and who adopts it? *Industrial and Labor Relations Review*. 47(2): 173–188.

Rajan RG. 2012. The Corporation in Finance. *Journal of Finance*. Blackwell Publishing Inc 67(4): 1173–1217.

Rajan RG, Zingales L. 1998. Financial Dependence and Growth. *American Economic Review* 88(3): 559–586.

Villalonga B, Amit R. 2006. How do family ownership, control and management affect firm value? *Journal of Financial Economics* 80(2): 385–417.

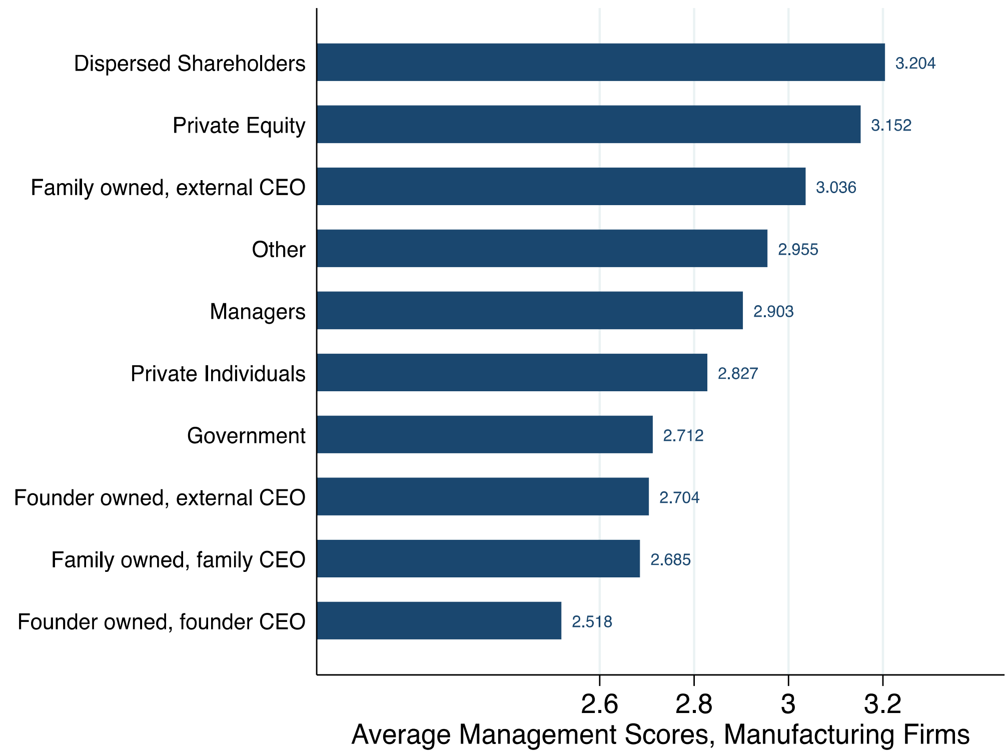
Wasserman N. 2003. Founder-CEO Succession and the Paradox of Entrepreneurial Success. *Organization Science*. 14(2): 149–172.

Wasserman N. 2006. Stewards, Agents, and the Founder Discount: Executive Compensation in New Ventures. *Academy of Management Journal* 49(5): 960–976.

World Values Survey Association. 2008. *World values survey*. Ann Arbor: University of Michigan.

APPENDIX

APPENDIX FIGURE 1
Management Scores Across Ownership Types



APPENDIX TABLE 1
World Management Survey (WMS) Questions

Practices	What we are measuring
Operations Management and Performance Monitoring	
Introducing Lean (modern) Techniques	Measures how well lean (modern) manufacturing management techniques have been introduced
Rationale for introducing Lean (modern) Techniques	Measures the motivation/impetus behind changes to the operational processes, and whether a change story was well communicated, turning into company culture
Continuous Improvement	Measures attitudes towards process documentation and continuous improvement
Performance Tracking	Measures whether firm performance is measured with the right methods and frequency
Performance Review	Measures whether performance is reviewed with appropriate frequency and follow-up

Performance Dialogue	Measures the quality of review conversations
Consequence Management	Measures whether differing levels of firm performance (not personal but plan/process based) lead to different consequences

Practices	What we are measuring
Target Setting	
Target Balance	Measures whether targets cover a sufficiently broad set of metrics and whether financial and non-financial targets are balances
Target Interconnection	Measures whether targets are tied to the organization's objectives and how well they cascade down the organization
Time Horizon of Targets	Measures whether the firm has a '3 horizons' approach to planning and targets
Target Stretch	Measures whether targets are based on a solid rationale and are appropriately difficult to achieve
Clarify and Comparability of Targets	Measures how easily understandable performance measures are and whether performance is openly communicated to staff

Practices	What we are measuring
Talent Management	
Managing Talent	Measures what emphasis is put on overall talent management within the organization
Rewarding High Performers	Measures whether there is a systematic approach to identifying good and bad performers and rewarding them proportionately
Removing Poor Performers	Measures how well the organization is able to deal with underperformers
Promoting High Performers	Measures whether promotion is performance-based and whether talent is developed within the organization
Retaining Talent	Measures whether the organization will go out of its way to keep top talent
Creating a Distinctive Employee Value Proposition	Measures the strength of the employee value proposition

Note: Survey Instruments with full set of questions asked are available at www.worldmanagementsurvey.org.

APPENDIX TABLE 2

Factors Correlated with Ownership Changes

	(1)	(2)	(3)
Dependent variable	Dummy = 1 if firm experiences a change in ownership between two survey waves, t and t+1		
Sample	All	Firms classified as Founder CEO at time t	Firms classified as different ownership at time t
Management Score (t)	-0.018	-0.055	-0.008
	(0.014)	(0.036)	(0.015)
ln(Firm employment) (t)	0.013	0.034	0.014
	(0.008)	(0.025)	(0.008)
ln(Firm age) (t)	-0.009	0.139**	-0.003
	(0.013)	(0.062)	(0.013)
Skills (t)	-0.006	0.022	-0.011
	(0.008)	(0.018)	(0.009)
MNE status (t)	0.008	0.229***	-0.009
	(0.019)	(0.070)	(0.020)
Constant	0.118	-0.052	0.024
	(0.116)	(0.334)	(0.145)
Observations	2844	493	2351
Adjusted R-Squared	0.131	0.143	0.156
Country dummies	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes
Noise	Yes	Yes	Yes

APPENDIX TABLE 3

Returns to Management for Different Strength of Family Values

	(1)	(2)	(3)
Dependent Variable	ln(sales)	ROCE	ROA
Family Values Index	-0.190*	-0.329	-362.284
	(0.104)	(3.626)	(242.689)
Management	0.077***	0.700*	27.214
	(0.019)	(0.392)	(25.130)
Family Values Index * Management	-0.023	-0.226	-9.652
	(0.026)	(0.532)	(33.069)
ln(firm employment)	0.636***	1.270**	54.603*
	(0.024)	(0.505)	(30.905)
ln(materials)	0.207***	0.859*	30.412
	(0.015)	(0.440)	(28.882)
ln(capital)	0.226***	-0.625	19.019
	(0.016)	(0.385)	(22.390)
Constant	3.108***	-14.062	155.132
	(0.290)	(9.580)	(836.400)
Observations	7,760	6,327	7,281
Adjusted R-Squared	0.808	0.084	0.080
Industry dummies	Yes	Yes	Yes
Firm employment	Yes	Yes	Yes
Firm age	Yes	Yes	Yes
Skills	Yes	Yes	Yes
MNE status	Yes	Yes	Yes
Noise	Yes	Yes	Yes

Notes: The sample used for this table includes only those firms for which sales, employment, capital, ROCE and ROA data could be found in ORBIS and other databases. All columns are estimated using OLS and standard errors clustered at the firm level. The FAMILY VALUES INDEX is taken from Bertrand and Schoar's (2006) survey of family values by country of CHQ location. *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

Essay 2 | Taking Stock of the Ability to Change: Prior Experience, Competency Traps, and Learning-by-Doing

Megan Lawrence

ABSTRACT

In aiming to understand why firms differ in their ability to adapt to new contexts, I explore the extent to which prior experience with a previous practice impacts the ability to properly execute a new organizational practice. I generate predictions by integrating the traditional organizational learning and path dependency literatures. To test my predictions, I use internal data from 294 stores of a large retail chain that implemented a new restocking process in its stores. Initial findings show that, on average, stores dramatically improve execution performance over time on average, and worker experience with the prior practice moderates the effect. Stores where employees have greater exposure to the old organizational practice perform significantly worse than other stores at the outset – consistent with the notion of 'competency traps'. However, these stores also learn more quickly, which I suggest may be the result of increased efficiency in their communications when learning the new process. These findings suggest that scholars in organizational learning should take attributes of an organization's experience history into account when assessing variations in new practice performance and learning.

I. INTRODUCTION

Sustaining a competitive advantage requires that firms respond to environmental changes by recombining and reconfiguring resources and organizational structures (Teece 2007). The advantage provided by a firm's current assets changes over time, most often eroding in value, so firms must undergo periodic efforts to redefine their current capabilities (Helfat 1994; Helfat and Peteraf 2003). Failure to navigate "competency-destroying" changes has been blamed for the death of companies like Kodak and Borders (Tushman and Anderson 1986; Rosenbloom and Christensen 1994). However, the adaptation and change required to maintain competitiveness in an industry frequently occurs at the daily activity or process level of the firm, requiring an incremental rather than radical strategic renewal (Agarwal and Helfat 2009). These changes in the individual and team level attributes of the business are subject to the same needs for adaptation and renewal to maintain competitiveness. Evolutionary theories of organization suggest that differences in certain attributes of the firms', and in these cases the teams', histories may cause some to adapt more quickly and some more slowly to these types of changes (Nelson and Winter 1982).

In this paper, I explore one determinant of how quickly teams adapt to properly executing a new organizational practice — the difference in prior experience with a previous, related practice. Two existing sets of theories lend themselves to predictions about the collective impact of previous worker experience on performance. The first set stems from the organizational learning literature. While that literature has broadly shown that learning from experience with a new technology improves performance over time (see reviews by Argote 2013; Dutton and Thomas 1984; Yelle 1979), the vast majority of learning-by-doing studies examine learning as if the technology is entirely new, not taking into account that most new technologies are actually variations on and replacements of old technologies. More recently, however, there has been some more specific work considering the relationship between prior experience and learning. For example, Schilling et al. (2003) consider whether related or unrelated variation in experience yields steeper learning curves in teams while Levitt, List, and Syverson (2013) examine the extent to which new automobile model variants and new shifts within manufacturing plants have similar learning curves to their contemporaries. Further, another literature on founders' prior work experience suggests that prior experience can be a source of knowledge that enhances the chances of survival for entrepreneurs in the personal computing (Bayus and Agarwal 2007), telecommunications (Chen, Williams, and Agarwal 2012), laser (Klepper and Sleeper 2005), medical device (Chatterji 2009), and financial services (Chatterji, de Figueiredo Jr, and Rawley 2014) industries. Combined, these literatures suggest that prior experience plays an important role for improvement in performance over time.

While the aforementioned works make predictions about improvement in performance, a second set of literature considers how the nature of path dependence and how the generation of competency traps within firms impacts performance more broadly. Increases in experience have been a source of rigidities in firms whereby performance suffers for some initial period (Leonard-Barton 1992; Levitt and March 1988; March 1991). Further, increased experience with a past practice engenders greater commitment to that practice and a greater integration of the practice into a system of interdependences that is not easily modified (Argote and Ingram 2000). Therefore, teams with different levels of experience with the prior practice may differ in their abilities to navigate the transition from the old to the new practice.

The learning and path dependence literatures have been compared before. Scholars have suggested that firms avoid costly competency traps is by employing 'slow learners' (Herriott, Levinthal, and March 1985; Lounamaa and March 1987; Denrell and March 2001; Rhee and Kim 2015).[16] By inserting these slow learners into an organization, the detrimental effect of commitment to an old practice can somehow be lessened for the overall group. However, integrating the literatures in this way neglects to consider that the incidence of competency traps may be jointly determined with the speed of the learners on a team or in a firm.

Considering these two theoretical perspectives differently, I develop integrated predictions about how differences in experience with a past version of a practice will affect both the performance and learning rate associated with the implementation of a similar but new practice. I test these predictions using a unique dataset constructed by gathering internal information from a Fortune 100 firm that was implementing a new restocking process in a subset of its retail stores. By capturing weekly data on this implementation and the attributes of the store more broadly, I am able to a) understand the raw performance each week and b) compute the extent to which learning occurs by store over time.

My findings indicate that, on average, stores improve significantly over the first few weeks of implementation. However, as predicted, improvement varies by store according to the amount of employee experience with previous practices. Consistent with the literature on competency traps, teams with higher proportions of team members with experience with the old practice perform worse initially. Consistent with the literature about learning, however, those teams also improve more quickly over time. Together these findings support the notion that differences in

[16] These 'slow learners' are individuals who learn and adapt to the organization more slowly than their peers.

the composition of team member experience, even across otherwise nearly identical settings and for relatively easy sets of tasks, have a meaningful impact on the performance of and improvement in new practice execution. The results also refute a misconception held by the strategy team at this retail firm's headquarters. The headquarters group believed that the teams with greater experience were "consistently underperforming and just not getting this new process." I suggest that the headquarters team may have evaluated the experienced teams too quickly and failed to consider the importance of improvement as a performance metric in new practice implementation in addition to the initial weeks' performance.

I consider several potential mechanisms to explain the results. Of all the mechanisms considered to explain the observed relationships among worker heterogeneity, team performance, and team learning, the only explanation consistent with all of the findings relates to teams learning communication over time (Zenger and Lawrence 1989) and creating mutual truces (Nelson and Winter 1982). This learned communication both inhibits initial performance – because these teams have developed patterns optimized for the old way of doing things – and enables accelerated learning of the new practice – because these teams already know how to work with one another to complete tasks. Because the nature of the competency trap is jointly determined by the same mechanism that enables faster learning, it may be a misconception to suggest that "slow learning" is a solution to the competency trap problem. Lastly, these results are consistent with a suggestion by a supervisor of the restocking process in one of the stores that "[restocking the store] isn't rocket science. The employees that do these jobs come in expecting to do the same thing every night, so a change to the routine necessarily disrupts the existing habits and quirks of each team."

This paper contributes to the current understanding of why firms may differ in their ability to recombine and reconfigure their capabilities. I am able to reconcile the notion that prior experience with related tasks can help organizations learn faster with the notion that experience with old practices might lead to competency traps. Because prior experience impacts both the learning speed and the existence of competency traps, a one-time adjustment to the learning speed of any team may not be a permanent solution to avoid competency traps among employees completing organizational practices. My theoretical resolution offers a micro-foundation for the existence of competency traps and evidence that they may exist even in something as seemingly basic as a restocking process. Empirically, I provide an alternative framework to evaluate learning rates such that I do not rely on the commonly held, but restrictive, assumption that the learning rate remains constant across time for a particular new practice implementation.

Because learning is critical to performance (Argote, Beckman, and Epple 1990), especially within the context of learning to change practices (Teece, Pisano, and Shuen 1997), this paper contributes to the understanding of a core strategy question: why do we see persistent performance differences among seemingly similar enterprises (see reviews by Bartelsman and Doms 2000; Syverson 2011)?

II. LITERATURE and THEORY

II.1 ORGANIZATIONAL LEARNING and PRIOR EXPERIENCE

Across a wide variety of settings and using many different metrics for performance, research in the organizational learning literature has shown that firms improve performance with repetition of a new task. Most early research associated with firm learning focused on the extent to which manufacturing firms exhibit ‘learning curves’ with costs declining with increasing cumulative output. Learning curves have since been estimated in both manufacturing (Argote, Beckman, and Epple 1990; Benkard 2000; Wright 1936) and services (Pisano, Bohmer, and Edmondson 2001; Reagans, Argote, and Brooks 2005; Darr, Argote, and Epple 1995) settings. Outcomes for learning studies have ranged from labor hours needed (Argote and Epple 1990) and customer satisfaction (Lapré and Tsikriktsis 2006) to defect density (Hatch and Mowery 1998) and survival rates (Baum and Ingram 1998). Many studies have focused on the determinants and moderators of these observed learning curves.

A persistent finding in this literature has been the large variability in learning rates across firms (e.g. Dutton and Thomas 1984). From an economic opportunity standpoint, this finding is puzzling since productivity differences are exceedingly large even within seemingly similar businesses (Syverson 2011). One of the many ways that these differences have been explained is through noting differences in the human capital involved in the production and / or execution of the new technology. While the knowledge present in a firm may reside in many places, individuals and teams are a fundamental component of any learning process (Grant 1996; Hitt et al. 2001; Felin et al. 2012). Therefore differences in individuals and teams which may affect learning can lead to these large variations in performance.

Because employees and teams learn with repetition, there is reason to believe that not just the experience with the new task but also the previous experiences of employees may affect firm performance and learning of a new task. After all, these new technologies in firms are often variations on old technologies and teams of workers implementing new technologies often have varying degrees of experience with the old technologies. The limited evidence that exists seems to suggest that prior experience is helpful. First, Pisano (1994) considers which types of tasks require learning-by-doing versus can be improved significantly by learning-before-doing. This learning-before-doing concept suggests that some teams that are more readily able to learn the task prior to the first execution may perform better at the outset. Similarly, the concept of developing an 'absorptive capacity' in firms suggests that there are investments in experience that may improve the ability to learn and improve at a new task once an opportunity arises (Cohen and Levinthal 1990). Second, learning studies have considered contexts where a second shift of workers joins a production line or a model variation is introduced on a production line (Epple, Argote, and Murphy 1996; S. Levitt, List, and Syverson 2013). In these settings, it is often the case that the introduction of a second team, model, or other variation after some previous production has taken place does not experience the same initial high level of costs incurred on the first introduction. Third, Schilling et al. (2003) considers a related but different question: the extent to which specialization on a particular task, as compared to execution of a variety of tasks, affects learning. This research finds that experience with a related task improves performance more quickly than experience with an unrelated task when alternating tasks over time.

In a somewhat broader context, the value of past experience for entrepreneurs and even diversifying firms that are either starting new firms or working in new contexts has been definitively positive. Past experience acts as a source of knowledge and provides a base of resources and capabilities that is useful at the outset of new endeavors (Helfat and Lieberman 2002; Klepper and Simons 2000; Klepper and Sleeper 2005). Having this experience enhances survival initially (Bayus and Agarwal 2007) and helps firms overcome growth impediments (Chen, Williams, and Agarwal 2012). Lastly, prior experiences can sometimes provide opportunities for the experiential learning that informs subsequent career moves (Chatterji, de Figueiredo Jr, and Rawley 2014).

When specifically thinking about prior experience of teams together, teams that have experience with each other may exhibit expedited learning as a function of their team member overlap. Studies have shown that members of organizations accumulate tacit knowledge faster than the same individuals working alone due to the influence of proximity on knowledge transfer (Hansen 1999; Szulanski 1996). With increased time spent together, team members learn to coordinate their activity as who knows what and who knows how to execute various tasks becomes more defined (Edmondson et al. 2003). As a result, these types of teams with shared experiences are better able to accurately locate, effectively share, and use knowledge to reach the desired outcome of assigned tasks (Reagans, Argote, and Brooks 2005). In these cases, the shared experiences allow for a shared language that facilitates communication (Arrow 1974; Allen and Cohen 1969; Tushman 1978; Zenger and Lawrence 1989). Therefore, teams with an ability to coordinate and communicate more effectively are expected to learn to improve faster at a new task.

While none of this literature has specifically examined the impact of having prior experience with an outdated task on a team's ability to improve, taken together this body of work points to prior experience as a valuable asset for performance improvement when implementing a new technology.

II.2 PATH DEPENDENCE and PRIOR EXPERIENCE

Path dependence in firms is an attractive notion since it accounts for the persistence of organizational features over time. When thinking about how a firm responds to changes, Teece, Pisano, and Shuen (1997) suggest that the ways in which an organization responds to shifts in the business environment are tied to its business processes, market positions, and expansion paths. Processes, in particular, are shaped by current and past asset positions and are not necessarily built to adapt to discontinuities. In this context, the human capital assets of a location may impact its ability to execute the new process.

Teams of employees that undergo a change in the way in which they execute their activities must break the learned pattern response that allowed them to effectively execute their old processes (March and Simon 1958). In order to do so, individuals and teams must undergo the task of 'unlearning'(Hedberg 1981; Nystrom and Starbuck 1984; Klein 1989) or 'forgetting' (Argote and Epple 1990; Shafer, Nembhard, and Uzumeri 2001) when existing knowledge becomes obsolete. While breaking the old patterned response is costly for individuals and teams, neglecting to adapt to the new process and continuing in the old process is costly to overall performance.

Specifically in the context of considering the impact of prior employee experience, scholars believe that teams with increased experience working together may suffer from competency traps or core rigidities when faced with a necessary change in the way activities are accomplished in firms (Leonard-Barton 1992; B. Levitt and March 1988). In cases such as the one for this study, where the old process allows for more idiosyncratic preference on behalf of the employees, increased history together likely results in more 'truces' being formed between employees about the way work is divided and coordinated (Nelson and Winter 1982; Weber and Camerer 2003). Therefore, in the context of a change, teams with greater amounts of shared history will have greater difficulty in immediately adjusting to the requirements of a new task. The embeddedness of the old process contributes to an inertia whereby change becomes increasingly difficult. However, the path dependence literature gives little evidence to suggest how teams might improve over time after the initial period of adjustment or to suggest the duration of the effect of a competency trap on performance.

II.3 INTEGRATED VIEW OF PRIOR EXPERIENCE

The literatures on learning and path dependence suggest two related but separate effects of prior experience on new practice implementation. First, prior experience with an old practice increases teams' abilities to coordinate with one another to understand a task. Teams that possess a shared language about the old task in light of the new, related task can communicate and coordinate activities using that already developed language. This ability to understand one another suggests that teams with prior experience together may improve performance more rapidly over time. In other words, prior experience may increase the rate at which teams learn.

Second, prior experience with an old practices increases the likelihood that teams form idiosyncratic understandings or 'truces' about how work is completed. These unspoken understandings about 'the way things are done' on the team may create issues when a new practice is instituted. The old ways of doing things may no longer apply, yet employees may be stuck in the old ways. This effect of prior experience may result in initially worse performance as employees adjust to the new process. The competency trap experienced by these teams affects the initial level of performance, rather than the rate of performance.

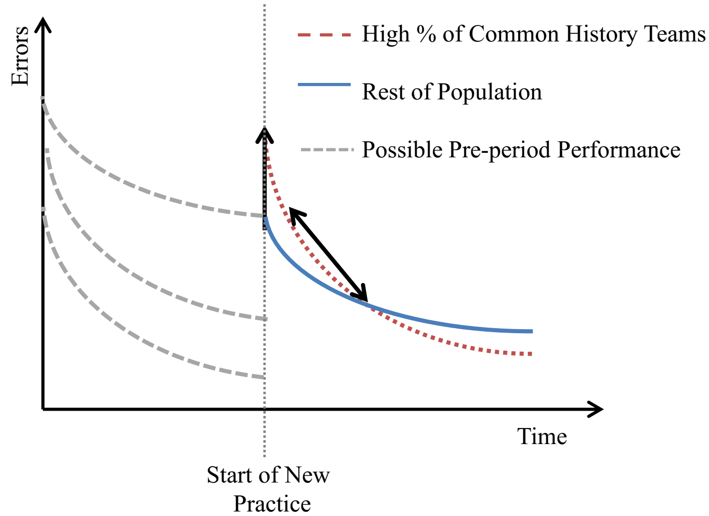
> *Hypothesis 1: Teams of employees with higher proportions of employees that have executed the old practice will learn more quickly.*

Hypothesis 2: Teams of employees with higher proportions of employees that have executed the old practice will perform worse immediately after a new process is implemented.

The joint prediction from both hypotheses is reproduced graphically in Figure 1 below. The grey dashed lines prior to the start of the new practice indicate a few potential pre-period performance paths. However, I make no inferences about the shape of the pre-period performance for specific teams. The predictions are generated only in the post-implementation period. The vertical black arrow indicates initially worse performance due to the competency trap experienced by these teams whereas the diagonal line indicates the rate of improvement, or learning rate, of the new practice. Teams with higher proportions of common experience with the prior practice, indicated by the red dotted line, will perform worse initially but improve more rapidly once the new practice is implemented.

FIGURE 1

Integrated Prediction for New Practice Performance: Teams with Higher Percentages of Common History versus Rest of Population



Notes: Grey dashed lines indicate examples of possible pre-period performance, but no predictions are made about this performance. The black vertical arrow at the start of the practice indicates worse performance initially as a result of changing practices while the diagonal line indicates a steeper learning curve once the new practice is in place. Teams with higher percentages of employees that have completed the old restocking practice will do worse initially but improve at a faster rate (red dotted line).

III. INSTITUTIONAL SETTING

In this study, I use data from a Fortune 100 retail firm making changes to its store restocking process in 2014. The firm is broadly classified as a big-box retailer and operates stores throughout North America. There is one restocking team per store, so the words 'store' and 'team' are used interchangeably.

Using a single firm, especially a retail firm, as a setting for a comparison study of its locations is advantageous for several reasons. First, by examining a new process implementation in a single firm across many locations, one can be confident that all stores were informed about the new process at the same time and were given a specific implementation plan for the change. This firm, in particular, operates thousands of locations and has standardized its stores such that within each one there is relatively little variation. Store layouts are standardized with regards to product availability and product placement. Similarly, the headquarters teams have outlined the specific tasks, timelines, and expectations for each employee role. When making a change to an aspect of the stores, every effort is made at the headquarters level to implement the change in exactly the same way across stores. Second, within a highly standardized firm such as this one, there is a similar, if not identical, culture and set of norms for operations across all locations. Of particular importance for this study, all candidate employees for each store are first screened through the headquarters before being allocated to specific stores that have a need for a new employee. Therefore, while the supply of employees may certainly differ from place to place, the process for hiring and training each employee is similar. Third, because of relatively low margins and the ever-present threat of losing market share to its competitors, this retailer, like many in the retail industry, keeps very detailed data on each of its stores on a daily and weekly basis in order to frequently assess the performance of each store. In summary, the study of a multi-unit retail firm is useful because stores are nearly identical with an ability to capture data for the ways in which they differ.

This firm made a change to its restocking process, specifically. Restocking a store generally includes three major functions: "receiving", "packout", and "packdown". Receiving encompasses all activities related to unpacking the trucks that deliver products. These trucks deliver approximately once a day, but some high-volume stores receive deliveries twice a day and some lower-volume stores receive less frequently. Within the truck, products are organized onto standard pallets (40" by 48"), which can be stocked up to approximately eight feet high with products. Usually, identical products will be on the same pallet; however the items on each pallet are arranged according to size and weight to ensure balanced pallets, so similar items may be separated in some instances. The receiving process involves unloading, unpacking, and sorting the shipment onto rolling carts that are assigned to specific aisles. For example, one cart may have products for aisles 1 and 2 while the second cart has inventory for aisles 3 and 4 and so on. The packout process involves moving products from the storeroom to the sales floor. After the receiving employees unpack inventory onto carts, the packout employees roll these carts to the appropriate aisles on the sales floor and pack the individual products onto the shelves. Any remaining inventory that does not fit on the shelves is left for the packdown. The packdown process entails moving products from storage shelves above reach in a given aisle to customer level when there is available space on the shelves as well as moving any extra inventory from the packout onto the shelves above for later use.

The new restocking process revised the current restocking process in three key ways. First, the number of departments supporting the restocking effort was consolidated from three to one in an effort to increase labor efficiency in restocking. The previous three departments aligned directly with receiving, packout, and packdown. In the old three-department organization, the exact timing needed for each of the functions had to be predicted accurately in order to avoid employee idleness – if a particular function took less time than usual, or overtime cost – if the function took longer than scheduled. In the new consolidated department, restocking employees can shift from receiving to packout or packout to packdown as needed. Second, the packout and packdown portions of restocking shifted from a system in which employees could packout or packdown any aisle, in any order, based on preference to a mandatory aisle-by-aisle, collaborative, shelf packout and subsequent packdown. Instead of an employee wheeling a cart filled with product to a specific aisle to complete packout, a set of employees, deemed ‘bowlers’, move the cart to the respective aisle and “bowl” the product down to its approximate area on the aisle. Then, employees come through behind the bowlers to put specific items onto shelves. The goal of making this process collaborative is to increase the positional accuracy of each item and, more importantly, to ensure items are not overlooked in the restocking process. Third, the hours of the packout and packdown employees were shifted to completely overnight rather than partially during store hours and partially overnight. By shifting the process overnight, the firm hoped to offer better customer service during the hours that the store is open and to reduce the disruption caused by closing an aisle to restock during the day. At the end of this entire process of receiving, packout, and packdown, a single employee has to certify all out of stock items by scanning shelf placements where the product is out of stock. This certification involves using a radio-frequency identification (RFID) handheld scanner to scan every empty product placement. This scan is matched with the computer-listed inventory of the item to certify items that are out-of-stock and re-order them from the regional distribution center.

In order to implement this change within stores, the headquarters provided the manager at each store with timelines, technical manuals, and video tutorials. The wide variety of tools were provided to guarantee that all stores received exactly the same information and planned to achieve the appropriate milestones associated with all practice changes. The headquarters also wanted to confirm, to the extent possible, that all information was translated uniformly. To achieve this goal, they included talking points in the provided materials catered to the type of meeting and type of employee for use during individual conversations and provided exact dates for certain conversations and announcements. The new process was announced a month and a half before it was implemented in the stores in order to allow time for management and employee adjustments. In the intervening time between the announcement of the change and the start of the new process, managers and employees were selected for the new consolidated restocking team as well as trained via informational packets and video demos. All employees – new and seasoned – were trained using the same materials and content. While the process of using the RFID scanner to close out the night does require an initial explanation, the store manager suggested that the process of scanning is incredibly simple, noting that "my pre-schooler could use these scanners after I showed him the right button." Also, the scanner is the universal tool used in the store, so at a minimum, the department manager in each store was already capable of using the tool to scan items at the outset of the process.

Because of the shift in hours of the main restocking roles to completely overnight, employees were given the opportunity to be placed into other departments at the store rather than to have to begin working overnight. Based on the need of the restocking department for additional employees after placing current restocking employees, employees from other departments in the store were allowed to indicate interest to switch onto the restocking team. Lastly, new employees were hired on a continuing basis since the attrition rate in retail establishments is perennially high. Therefore, as a result of this restocking change, each store had up to three types of employees on a team executing the new restocking practice. First, there are those who were in one of the three restocking departments of the store prior to the change; I call these employees "department veterans." Second, there are those employees who were in another department but opted to switch into the restocking department at the time of the change. I call these employees the "transfer veterans." Third, there are "new employees" – employees who recently joined the firm. The proportions of each employee type vary widely across stores.

The process by which restocking at this firm was changed is also advantageous for my study. This restocking change was tested across some small sub-populations of stores within the firm before being implemented more widely across the grouping of stores which I analyze for this study. Because the exact change to the restocking department and process involved reducing the number of department managers in line with the reduction in departments, the change was not widely publicized on the firm intra-web as would have been the norm for the rest of the store initiatives issued from the headquarters strategy group. It was shared privately only to the subset of stores and store districts that implemented the process in the given time period. This lack of published communication about the implementation reduces concern that employee behavior and / or store performance would be different in anticipation of the change for the departments that I observe. Also, since the restocking process is completed overnight, these employees have no other task besides restocking. Because the store is not open and there are no customers to serve, any time that is not spent in an effort to restock is not spent productively. Further, there were no formal incentives associated with staying in the restocking department, switching into the department, or switching out of the department as a result of this change. Employees who chose to leave or join the department at the time of this change were offered their current salary in whatever new role they took on (with the exception of employees who were promoted to management). Lastly, all employees, new hires and current employees, received the same training on the new restocking method.

More broadly, a study of a restocking process change is advantageous for study purposes since it is a process that has proven challenging for companies to change as evidenced by numerous popular press articles about failed restocking experimentation in the media (e.g. Moylan 2015; Saporito 2013). Experiments with the dynamic management of employees in retail have even resulted in product buildup in stockrooms without appropriate personnel to bring products to shelves for purchase. While some of these retailer experiments may have been extreme cases, there are real costs associated with misplaced and out-of-stock items including missed sales and lost customers (Fleisch and Tellkamp 2005; Kang and Gershwin 2005; DeHoratius, Mersereau, and Schrage 2008). Further, as a dynamic process itself, restocking is a practice that requires adjustment over time as patterns of shipment and sales change inventory levels and new technologies become available for both customers and retailers.

IV. DATA and EMPIRICAL STRATEGY

IV.1 DATA

The data for this study were compiled from more than 400[17] internal reports from various departments within the focal retail firm. Data from 294 stores located in the Northeast and Midwest regions of the United States that implemented the new restocking process in early 2014 are included in this study. In order to compile the data, I first aggregated the weekly

[17] Reports for each district of stores for each week were stored separately. All additional information pertaining to store locations, attributes, sales, and employment were also kept in separate reports.

performance data directly from archived reports produced weekly for every district of stores. Then, I matched this store-week performance with employee information from human resources, with historical information on each store, and with weekly data related to sales for each store. For this analysis, I use data from weeks 2 through 12 of the implementation of the new process in the stores. Data from week 1 of the implementation were not collected by headquarters.

IV.1.1 DEPENDENT VARIABLE

The dependent variable used for this analysis is a metric called a "missed scan." It is a count of the total errors made each week in the end-of-night reconciliation and certification component of the process. More specifically, at the end of each night of restocking, one employee per store is responsible for examining each aisle and using a handheld RFID scanner to scan all shelf placements for which the appropriate SKU-specific product appears to be missing. A missed scan results from a SKU-specific product showing as out-of-stock in the store's computer records and its appropriate location in the store not being scanned to indicate an out-of-stock. Because inventory records are often inaccurate (DeHoratius and Raman 2008), this reconciliation process of matching computer-listed out-of-stock items with their empty shelf placements is the method by which items are reordered from the regional distribution center. This reconciliation scanning is the latest way that the firm has tried to deal with identifying out-of-stock products in a timely fashion while not holding extra inventory in stores. While the scanning requires an extra step in the restocking process, rather than just reordering based on the computer records, headquarters considers this step useful to the overall process of maintaining accurate stock counts and measuring restocking performance.

There are several store scenarios that may result in a missed scan in a given week. I will describe the most common reason, as was described by a store manager in one of the affected stores. Imagine that a customer picks up a product in order to decide whether or not to purchase it. The customer carries the item around in his basket before subsequently deciding not to purchase the item. He removes the item from his basket and places it on an empty spot on the shelf (assume that the product in this location is out-of-stock so the placement location is empty). In the event that this customer returned the item to a shelf placement other than its appropriate location from which it was first picked up, the product now occupies a placement location for an item that is out-of-stock. If, in the course of the restocking process that night, the affected item is not removed from the erroneous product location and restocked in its appropriate location, the SKU-specific product for the location that the focal product is occupying will not be appropriately identified as out-of-stock and scanned, which will result in a missed scan.

A second main reason that a missed scan can occur is if the restocking team does not complete its packout and packdown of the store prior to the appointed time that the employee scans the shelves. This miscommunication can happen if the employee appointed as the scanner does not know that the packout and packdown of an area has not been completed. For example, if the packdown team has to skip an aisle and plans to come back to it later but does not notify the scanner, the scanner may scan an aisle prematurely. This premature scanning can present an issue because employees are responsible for tidying and replacing misplaced product in the shelves in addition to restocking the shelves. If they do not complete their aisle work, the scanner is much more likely to miss product locations when scanning out-of-stock items. In either of these scenarios, a missed scan is problematic because it results in a product not being reordered from the district distribution center and persisting as a costly out-of-stock.

This missed scan metric is an appropriate measure of process performance because it captures a measure of the restocking team's ability to execute on the goals of the restocking process. First, the restocking team is responsible for packing down all available products to their appropriate locations on the shelves. Second, the restocking team is responsible for organizing and tidying all shelves and product locations which includes returning all misplaced items to their appropriate SKU-specific product locations. Missed scans result from a failure to appropriately execute on these two tasks. Because the retailer relies on the in-stock number for each SKU being accurate in order to reorder items, this process of checking the out-of-stock items by scanning ensures that the right products are being reordered for each store. Lastly, these missed scans are seen by the headquarters management team as an error by the restocking department team and are sought to be minimized over time. They are a measure of performance upon which these restocking teams are evaluated on a weekly basis.

IV.1.2 INDEPENDENT VARIABLES

IV.1.2.1 New Process Experience

The main variable of interest in understanding the overall learning that occurs as this new process change is implemented is the experience each store has in executing the new process. The restocking practice happens each weeknight - meaning a total of five times per week. Therefore, as a measure of experience, I use the week of implementation as a proxy for experience. For example, at the conclusion of week 2, each store has performed the new restocking process 10 times. This weekly measure is both convenient, since the other variables of interest are often measured weekly, and practical, since a weekly measure allows for a level of smoothing across differences in restocking need due to the day of the week. This approximation does not suffer from the normal shortfalls of time-indicators in the learning literature since it is an exact replacement for the number of completed rounds of the new process, analogous to the production volume measures used in other models (Darr, Argote, and Epple 1995; Lieberman 1984; Rapping 1965).

In this analysis specifically, I introduce an indicator variable for each implementation week starting with week 3 and continuing through week 12. The omitted week is implementation week 2, so estimated coefficients on the week indicators reflect the change in missed scans relative to the second week of implementation. This method of measuring experience using indicator variables rather than a single measure allows for a flexible learning rate, which I discuss in more depth in the Empirical Strategy section.

IV.1.2.2 Prior Process Experience Measures

Proportion of Department Veterans: In each store, there are different proportions of employees that are so called "department veterans," or those employees that have direct experience executing the old version of the restocking practice. In order to test whether these employee types differentially impact performance, I create a measure of the proportion of each store's employees that are department veterans each week. This variable ranges from 0 to 1 depending on the composition of the restocking teams.

Proportion of Transfer Veterans: I create a measure of the proportion of each store's employees that are "transfer veterans" each week. Transfer veterans are those employees that were working for the firm previously in a different capacity and moved into the restocking department at the time of the process change. These employees have indirect experience with the old process through conversations and general knowledge about the store environment and direct experience with product placements and other aspects of employment (e.g. employee code of conduct, expectations of employment).

Proportion of Old Guard: I create a measure of the proportion of so called “Old Guard” employees on each restocking team. This measure is equal to the sum of the proportion of department veterans and the proportion of the transfer veterans. These employees have experience, either directly or indirectly, with the old way of doing things. They share a common language about ‘the way things used to work’ which is important to consider when examining the method by which employees improve at the task.

IV.1.3 CONTROLS

Weekly Sales Volume: The number of missed scans per week might be associated with the number of products that are moving around on shelves each week. One proxy for the extent of the restocking work to be completed is the weekly sales volume. For example, in the extreme case where no one enters a store and nothing is purchased in a particular store in a given week, the store will likely miss very few scans since very few, if any, items moved locations. In contrast, a store that does an exceptionally high amount of volume in a given week may face much more difficulty accurately completing the restocking task each night. By accounting for store-specific sales per week, I account for differences associated with any seasonal or idiosyncratic increases or decreases in sales volume.

Average Ticket Size / Average Units per Transaction: I use two measures to account for potential differences in the purchasing behavior of customers. Purchasing behavior differences may result from any number of different attributes of a store-week such as seasonal or holiday versus non-holiday purchasing. Different purchasing behaviors may result in differences in the type and number of products purchased in each store. While the exact effect of any of these singular differences on composition of products purchased is beyond the scope of this study, I use two proxy measures to account for broad differences in the purchasing behaviors as related to the effect on the number of missed scans. First, I control for the average ticket size in a given week. The ticket size is defined as the dollar value of a transaction; therefore, the average ticket size is the average value of each transaction in a given store in a given week. Second, I control for the average units per transaction. The units per transaction are the number of items purchased in a single transaction, so the average units per transaction variable captures the average number of products purchased per transaction for a given store in a given week.

Average Employee Experience: By compiling data on the start dates of each employee on the restocking team in each store, I am able to calculate an average experience of the restocking teams by store-week. This measure is calculated in years. By allowing this measure to accumulate throughout the progressive weeks of the implementation, I am implicitly accounting for entry and exit of employees in various weeks. By controlling for experience of the restocking team broadly, I account for any pure tenure-related differences that may have an effect on teams ability to complete the process.

Employee Experience Shannon: Even though two restocking departments may have identical average employee experience measures, they may have very different distributions of employee experience on their teams. Therefore, I create a measure of experience concentration using the experience levels of each employee in a given restocking department. As a retail firm, there is often high turnover, resulting in many employees with low experience; however restocking teams at this firm in particular often have extremely experienced members as well. The experience concentration value for each store-week is calculated as a Shannon index indicating the extent to which the accumulated experience of the team in aggregate is concentrated in a few members of the team.[18] The Shannon index takes values greater than or equal to zero with the value rising as experience is more evenly spread amongst the members of the team. This measure has often be used in diversity constructs to operationalize variety (e.g. Harrison and Klein 2007) and in this case is indicative of the evenness of the experience accumulation in the department.

Supervisor Experience: The restocking team at each store has a supervisor, or manager, who oversees the tasks of the group, schedules employees, and reviews team and individual performance. While the restocking employees' experience levels are the main focus of this study, the supervisor's experience at the firm may also impact the ability of the teams to learn the new process. The supervisor of the restocking department often has many years of experience within the firm; however the supervisor is not necessarily the most senior person in the restocking department. The supervisor experience variable is measured in years of experience.

Employee Count: Employee count is a measure of the number of employees on a store's restocking team in a given week. The variable takes values for the total number of employees that were employed for any amount of time during the focal week. Because this particular firm was implementing the new restocking process as it entered its busy season, the number of

---

[18] Shannon Index $= -\sum_{i=1}^{n}[p_i * \ln(p_i)]$ where p is the proportion of the total team experience (in years) attributed to employee i. The total number of employees on the restocking team in a given store is noted by n. As an example, if there were 2 employees and one employee had 1 year of experience while the other employee had 2 years of experience, the Shannon Index would be calculated as follows:

$$= -\left(\frac{1}{3}\right) * \ln\left(\frac{1}{3}\right) + -\left(\frac{2}{3}\right) * \ln\left(\frac{2}{3}\right) = 0.6365.$$

employees in the department was expanding over the time period to account for the increased restocking needs.

New Employee: The new employee entry variable counts any new employee with a start day between the end of the prior week and the end of the current week as a new employee, regardless of whether the employee started at the beginning or end of the week.

Employee Exit: Similarly, I count the number of employees who leave work at the store in a given store-week in the employee exit variable. This exit variable captures employees who left voluntarily and involuntarily and includes all employee exits between the end of the prior week and the end of the current week.

Fixed Effects: In order to account for any time-invariant differences between stores that may bias estimates on the human capital variables of interest, I use fixed effects at the store level. While I chose a setting for study where stores are highly standardized, there are still ways in which the stores may fundamentally differ from one another in ways that affect performance. Fixed effects limit the ability to make observations about the impact of these fundamental differences by eliminating any variation at the store level. For example, if performance differs by store because some stores are in urban locations and other stores are in suburban locations, that difference in performance will be absorbed by the store fixed effect as will other similarly time-invariant differences between the stores. However, the advantage of using fixed effects in this analysis is that it allows for a more focused study of exactly the impact of sharing a common history with an old practice on same store performance and learning while eliminating differences stemming from more systematic differences between the stores.

IV.2 EMPIRICAL STRATEGY

The traditional learning curve function models a pattern of cost reduction or other improvement with the form

$$C(x) = ax^{-\beta} \tag{1}$$

where x is experience with the new practice or technology, $C(x)$ is the unit variable cost or performance after x units of experience, a is the starting cost, and β is the constant learning parameter. This model assumes a constant learning rate over time. While there may be periods of relatively consistent improvement, in many situations a plateau of performance is reached whereby the learning rate necessarily drops (Thompson 2012; Yelle 1979).

Therefore, rather than a traditional exponential functional form, which restricts learning to a constant rate, my model of learning-by-doing allows for a non-constant learning rate. In order to create a model that accommodates these changes in learning rates, I include an indicator variable by week excluding the initial week. By omitting the indicator variable for the first week in my sample, all future week variable coefficients are estimated in reference to the initial week. The base specification is as follows

$$Y_{it} = \alpha + \beta_1(week3_i) + \beta_2(week4_i) + \cdots + \beta_9(week11_i) + \beta_{10}(week12_i) + \gamma_i + \varepsilon_{it} \tag{2}$$

where α is a constant and γ_i is the store fixed effect. The week variables are equal to 1 for the focal implementation week and 0 otherwise. The coefficients β_j, where $j = 1,2,..10$, are estimated as the average decrease in missed scans from the first week of the data. This nonparametric method of measuring learning is able to accommodate the often sudden flattening out of the learning curve by allowing for a varying learning rate over time.

The store level controls associated with sales are added to equation (2). Then, the prior process experience variables are added, first testing the raw percentage of the team around before the change (*% Old Guard*) before examining the team breakdown more specifically. The general specification for these models is as follows

$$Y_{it} = \alpha + \beta_1(week3_i) + \beta_2(week4_i) + \cdots + \beta_9(week11_i) + \beta_{10}(week12_i) + \sum_{j=1}^{m} \delta_j controls_i + \sum_{k=1}^{n} \rho_k(PriorProcessExperience\ Var_i) + \gamma_i + \varepsilon_{it} \quad (3)$$

where estimates on the prior process experience variables can be interpreted as differences in levels of performance associated with these variable values in the specified weeks.

The specifications so far are useful for understanding how raw performance varies based on differences in exposure to the prior process, but these differences may also affect the rate at which teams improve at the new task. In traditional learning models, differences in learning rates are measured by interacting the constant learning rate with an additional variable of interest. In this case, the rate is captured by interpreting the coefficients on 10 new process experience variables so interacting these variables with the variable of interest would be both tedious and difficult to interpret.

Instead, I modify the model to fit a non-linear piecewise function with two segments to the data to test for differences in learning rates. The resulting spline has two variables measuring the new process experience weeks that can be interacted with the prior process experience variables of interest to capture differences in learning rates. The specification is as follows

$$Y_{it} = \alpha + \beta_1 week_i + \beta_2 week2var_i + \sum_{j=1}^{m} \delta_j controls_i + \sum_{k=1}^{n} \rho_k(PriorProcessExperience\ Var_i) + \vartheta_1(week_i * \%Old\ Guard_i) + \vartheta_2(week2var_i * \%Old\ Guard_i) + \gamma_i + \varepsilon_{it} \quad (4)$$

for the case of interacting the week variables with the *% Old Guard* variable. The first week variable, *week*, takes the values of the week of implementation while the second week variable, *week2var*, takes a value of 0 for the first segment of the piecewise function and takes a value of the implementation week minus the week in which the break occurs for the second portion of the nonlinear function. I determine the break for the piecewise function using a nonlinear least squares estimation. If teams with higher proportions of employees that have experience with the old way of doing things learn more quickly, this specification will reflect the increased learning through a negative and significant coefficient on the *%Old Guard*wk* interaction

V. RESULTS

Using the data from the 294 retail stores for weeks 2 through 12 of the implementation of the new restocking process, I compile 2,872 store-week observations for use in this analysis. From the summary statistics in Table 1 below, note that the dependent variable, missed scans differs dramatically. The average missed scan in a given week is equal to about 18.9 while the highest missed scan overall is 295 missed scans (in week two by one store) and the lowest missed scan is 0 (by several stores in the later weeks of observation). For the purposes of this analysis, I omitted observations for which the missed scans were greater than 400 scans. These observations all occurred in the first week of the data and seemed to be a result of overall misunderstanding about the process rather than actual attempts at the correct new process. The extraordinary circumstances of these first week errors were confirmed by asking a store manager how this many errors would have occurred on the first week. The store manager noted that errors of this magnitude in the first week could have only occurred if "those stores must have not completed the scanning process nearly at all in the first week." Later, in the robustness analysis, I test alternative ways of including high missed scan values to confirm that the main results hold.

Table 1 also includes attributes of the stores in general as well as characteristics of the human capital for each store-week. The average store is about 13.1 years old and about 109,000 square feet. The average store sells about $34.3MM of merchandise per year, with weekly sales in this period of time netting about $722,000 a week.

TABLE 1

Summary Statistics

Variable	Observations	Mean	Std.Dev.	Min	Max
Missed Scan	2872	18.915	24.828	0	295
Weekly Sales Volume	2872	722,364	271,888	217,509	1,874,356
Yearly Sales Volume	2872	34,400,000	11,500,000	16,100,000	85,800,000
Store Age	2872	13.168	4.008	2.32	25.23
Store Square Feet	2872	108,616	10,473	62,719	138,975
Average Ticket Retail Value	2872	61.612	9.550	35.390	117.266
Average Units per Transaction	2872	6.342	0.669	4.320	10.677
Average Employee Experience (yrs)	2872	3.274	1.487	0.714	10.497
Supervisor Experience (yrs)	2872	7.867	4.781	0	26.808
% of Department Veterans	2872	44.8%	0.152	9.1%	100.0%
% of Transfers	2872	8.3%	0.074	0.0%	46.7%
% of New-ish Employees	2872	46.9%	0.157	0.0%	85.7%
Experience Shannon	2872	4.574	0.711	2.762	7.917
Employee Count	2872	19.762	4.885	7	39
New Employee	2872	0.594	0.956	0	6
Employee Exit	2872	0.399	0.707	0	7

The mobility and experience of employees differs widely by store-week. The average experience of a team member in the restocking department, even taking into account the new employees, is about 3.3 years. This average may seem quite high; however this firm, in particular, has been known for having retail employees with upwards of 20 years of experience at their stores. Supervisors usually, but not always, have much more experience than employees on their teams. Their overall average experience is 7.86 years. After the change in the restocking process, teams are mostly composed of either employees that worked on the old restocking process (44.8% of employees) or newer employees who joined after the restocking change was announced (46.9% of employees), but there are about 8% of employees that transferred into the restocking department from another department at the store.

The team sizes range from 7 to 39 employees, but there is an average of about 20 employees per store-team. The employee team composition often fluctuates as approximately 2 new employees show up every 3 weeks, resulting in an average of about 0.59 new employees per week. The maximum number of employee exits in a week was 7; however that number is an outlier since the average employee exit per week is about 0.40 employees.

In examining the correlation matrix in Table 2, trends emerge in store type. Stores with higher sales volume tend to be older and larger. Also, stores with higher average ticket sales have higher average units sold per transaction, perhaps unsurprisingly. Mostly, it is important to note that none of the variables capturing the firm experience and / or old process experience are highly correlated.

TABLE 2

Correlation Coefficients

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1)	Missed Scan	1															
(2)	Weekly Sales Volume	-0.1585	1														
(3)	Yearly Sales Volume	0.0171	0.8326	1													
(4)	Store Age	0.0396	0.374	0.431	1												
(5)	Store Square Feet	0.0787	0.2676	0.2991	0.3003	1											
(6)	Average Ticket Retail Value	-0.0202	-0.0996	-0.1094	-0.1604	-0.1273	1										
(7)	Average Units per Transaction	-0.1447	0.187	-0.0064	-0.1226	-0.0086	0.5144	1									
(8)	Average Employee Experience	0.0306	0.3665	0.4546	0.3502	0.19	-0.0991	0.0229	1								
(9)	Supervisor Experience	-0.0286	0.0545	0.0542	0.1082	0.0477	-0.0542	-0.0383	0.2983	1							
(10)	% of Department Veterans	0.084	0.246	0.3605	0.111	0.1598	-0.0705	-0.0335	0.674	0.1054	1						
(11)	% of Transfers	0.0963	0.1196	0.1697	0.0912	0.1243	-0.1018	0.0245	0.1548	-0.0685	-0.17	1					
(12)	% of New-ish Emp	-0.1263	-0.2939	-0.428	-0.15	-0.2127	0.1158	0.021	-0.7241	-0.0699	-0.8873	-0.3036	1				
(13)	Experience Shannon	0.1498	-0.0471	0.0207	0.049	0.0023	-0.0468	-0.0988	-0.3334	-0.0608	-0.3889	0.12	0.3199	1			
(14)	Employee Count	-0.0468	0.6087	0.676	0.3311	0.2875	-0.2873	-0.1009	0.0377	0.0481	-0.0024	0.0549	-0.0233	0.023	1		
(15)	New Employee	0.1054	-0.0638	0.0035	0.0368	0.0071	-0.037	-0.0464	-0.1188	-0.0045	-0.1173	-0.0496	0.1366	0.1105	0.1245	1	
(16)	Employee Exit	-0.0248	0.0286	0.0114	0.0309	-0.0033	-0.0334	-0.008	-0.0904	-0.0267	-0.1025	0.0209	0.0893	0.2901	0.0224	0.062	1

Examining both the summary statistics and the correlation coefficients provides a general validation of my choice to include store fixed effects. While stores do not necessarily vary as drastically as if comparing a mom-and-pop store to a big box store, there is a fair amount of variation in the size of these stores. Additionally, the staffing of some stores is multiples larger than others, which means that the coordination effort and familiarity of workers is very different between the top and bottom end of the store size range. The fixed effect specification adds a control at the store level such that all the non-varying aspects of the stores that may differ will not bias the findings.

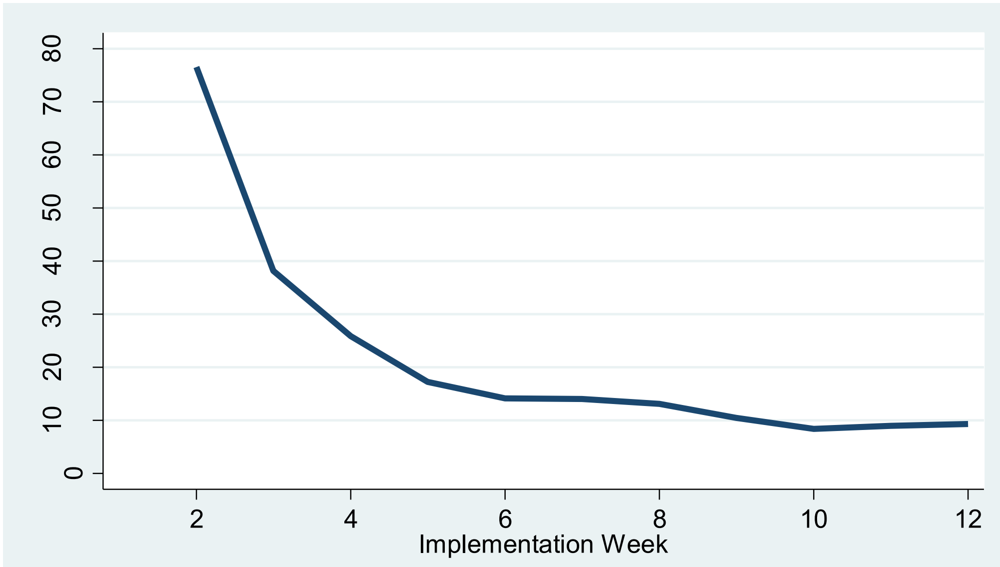
V.1 LEARNING TO REDUCE ERRORS

In Table 3, I provide baseline estimates of comparison, taking into account the store-specific attributes. From Specification (1) in Table 3, it is clear that the stores make fairly dramatic improvements in the first three weeks before making more incremental improvements in later weeks. From week 2 to week 3, the average store misses 31.1 fewer scans and from week 3 to week 4, the average store misses 12.2 fewer scans and so on.[19] From just the raw weekly coefficients in Specification (1), it is important to note that the learning rate is not constant. The learning rate can be calculated by noting the improvement week-over-week as compared to the weekly baseline. The learning rate is quite high at first, on the order of 45% improvement (1-(69.243-31.062)/69.243), but lowers to less than 5% by week 7 before picking up again around week 9. Therefore an exponential or power function, as is common in the learning literature, would incorrectly specify the learning rate by smoothing these different rates into an average improvement rate which underestimates the initial learning rate and overestimates the later learning rate. Another way to understand these coefficients is with a plot of the data as in Figure 2.

[19] The improvement week-over-week is calculated by subtracting the coefficients from adjacent weeks for all weeks after week 3. For example, from week 3 to week 4 the improvement is calculated as follows ($-43.296 - -31.062 = -12.234$).

FIGURE 2

Raw Average Missed Scans by Implementation Week



In Specifications (2) – (4) of Table 3, I add in the control variables which are related to the movement of the product on shelves and are suspected to impact the missed scan performance metric. General store effects are controlled for in the store fixed effect, but these measures provide indications of week-to-week differences in the customer segment in each store. In Specification (2), I add the weekly sales (in thousands). Missed scans are significantly positively impacted by the weekly sales of a store; however, each extra $100,000 in sales increases the number of missed scans by only 1.2 each week. Said another way, a one standard deviation increase in sales would only boost the missed scans by 2.52 points. In Specifications (3) and (4), I include the average ticket retail and average units per transaction measures. While these two measures are not significant in the baseline regressions, I keep the average units per transaction measure in future regressions since it provides an additional indication of the relative quantity of items that move locations in the store in a given week. After controlling for the fluctuations in store volume and customer behavior, the improvement in store performance week-over-week persists and is even slightly magnified.

TABLE 3
New Practice Learning in Stores

	(1)	(2)	(3)	(4)
	missedscan	missedscan	missedscan	missedscan
wk3	-31.062***	-31.119***	-31.421***	-31.296***
	(1.875)	(1.874)	(1.897)	(1.898)
wk4	-43.296***	-43.862***	-44.498***	-43.897***
	(1.877)	(1.888)	(1.989)	(2.022)
wk5	-51.739***	-53.522***	-54.211***	-53.357***
	(1.891)	(2.012)	(2.123)	(2.186)
wk6	-55.155***	-56.853***	-57.527***	-57.105***
	(1.884)	(1.994)	(2.102)	(2.117)
wk7	-55.303***	-56.930***	-57.858***	-57.754***
	(1.875)	(1.976)	(2.177)	(2.177)
wk8	-56.174***	-58.512***	-59.478***	-59.132***
	(1.870)	(2.076)	(2.285)	(2.294)
wk9	-58.907***	-62.054***	-63.206***	-62.701***
	(1.878)	(2.237)	(2.509)	(2.527)
wk10	-60.791***	-64.319***	-65.579***	-65.334***
	(1.881)	(2.323)	(2.635)	(2.639)
wk11	-60.240***	-64.286***	-65.621***	-65.426***
	(1.873)	(2.441)	(2.774)	(2.775)
wk12	-61.863***	-67.027***	-68.537***	-68.195***
	(2.242)	(3.003)	(3.352)	(3.358)
Weekly Sales (000s)		0.012***	0.014***	0.017***
		(0.004)	(0.005)	(0.005)
Avg Ticket Value			-0.107	-0.081
			(0.106)	(0.107)
Avg Units (per trans)				-1.785
				(1.095)
Constant	69.243***	63.068***	68.920***	76.266***
	(1.544)	(2.847)	(6.436)	(7.855)
Fixed Effect:	Store	Store	Store	Store
Obs.	2872	2872	2872	2872
R-Squared	0.526	0.527	0.527	0.527

Notes: Dependent variable is the missed scan errors. All columns estimated by using a fixed effect model at the store level. Therefore, results are reflective of the effects of within store variation on missed scan errors. The *wk#* variables take a value of 1 when it is the focal week and 0 otherwise (e.g. *wk3* = 1 for week 3 and 0 otherwise). The omitted week is week 2. *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

After examining the average relationship between overall store experience with the restocking process and performance, I control for the effect of raw firm-specific tenure of employees in

Table 4. In Specification (2), I find that increasing the average tenure of the employees on the restocking team increases the number of missed scans per week. I find no general impact of supervisor tenure on performance (Specification 3), but I include it as a control in future specifications since changes in the department leads were a key aspect of this process change. In Specification (4), I include a measure of the concentration of the team tenure. The coefficient on the *Experience Shannon* variable is not significant, meaning that the concentration of firm-specific tenure in a few members of the team as compared to more evenly distributed among the members does not affect the errors in the new process implementation. In other words, teams with a few very experienced people and the rest new employees do not perform significantly differently from teams where all employees are moderately experienced.

TABLE 4

Effect of Experience on New Practice Performance

	(1) missedscan	(2) missedscan	(3) missedscan	(4) missedscan	(5) missedscan	(6) missedscan
wk3	-31.065***	-30.768***	-30.765***	-30.770***	-30.099***	-30.068***
	(1.873)	(1.870)	(1.871)	(1.871)	(1.876)	(1.876)
wk4	-43.387***	-42.685***	-42.681***	-42.680***	-41.509***	-41.479***
	(1.906)	(1.912)	(1.912)	(1.912)	(1.935)	(1.936)
wk5	-52.790***	-51.749***	-51.743***	-51.733***	-50.115***	-50.114***
	(2.053)	(2.068)	(2.069)	(2.070)	(2.113)	(2.113)
wk6	-56.580***	-55.379***	-55.366***	-55.344***	-53.295***	-53.324***
	(1.999)	(2.022)	(2.023)	(2.026)	(2.100)	(2.100)
wk7	-57.064***	-55.776***	-55.759***	-55.731***	-53.353***	-53.383***
	(1.976)	(2.004)	(2.006)	(2.009)	(2.111)	(2.111)
wk8	-58.397***	-56.955***	-56.940***	-56.897***	-54.095***	-54.098***
	(2.076)	(2.110)	(2.112)	(2.118)	(2.251)	(2.251)
wk9	-61.819***	-60.249***	-60.231***	-60.176***	-57.035***	-57.029***
	(2.240)	(2.276)	(2.279)	(2.290)	(2.442)	(2.442)
wk10	-64.389***	-62.733***	-62.710***	-62.643***	-59.061***	-59.070***
	(2.323)	(2.362)	(2.365)	(2.380)	(2.570)	(2.570)
wk11	-64.430***	-62.840***	-62.814***	-62.747***	-58.912***	-58.919***
	(2.442)	(2.475)	(2.479)	(2.494)	(2.703)	(2.703)
wk12	-67.060***	-65.572***	-65.546***	-65.467***	-61.196***	-61.233***
	(3.002)	(3.023)	(3.027)	(3.043)	(3.255)	(3.255)
Weekly Sales (000s)	0.015***	0.015***	0.015***	0.015***	0.015***	0.015***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Avg Units (per trans)	-1.910*	-2.094*	-2.098*	-2.099*	-2.215**	-2.228**
	(1.082)	(1.080)	(1.081)	(1.081)	(1.079)	(1.079)
Emp.Experience (yr)		4.720***	4.738***	4.254*	-3.597	-3.415
		(1.303)	(1.307)	(2.308)	(2.666)	(2.671)
Supervisor Experience (yr)			-0.219	-0.191	0.037	0.078
			(1.196)	(1.201)	(1.195)	(1.196)
Experience Shannon				0.627		
				(2.459)		
% Old Guard					61.306***	
					(17.103)	
% Dept Veteran						54.750***
						(18.169)
% Transfer Veteran						92.793***
						(34.065)
Constant	72.478***	56.958***	58.615***	57.146***	50.150***	49.614***
	(6.041)	(7.394)	(11.697)	(13.040)	(11.907)	(11.917)
Fixed Effect:	Store	Store	Store	Store	Store	Store
Obs.	2872	2872	2872	2872	2872	2872
R-Squared	0.527	0.530	0.530	0.530	0.532	0.532

Notes: Dependent variable is the missed scan errors. All columns estimated by using a fixed effect model at the store level. The *wk#* variables take a value of 1 when it is the focal week and 0 otherwise (e.g. *wk3* = 1 for week 3 and 0 otherwise). The omitted week is week 2.*** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

V.2 PRIOR EXPERIENCE and INITIAL PERFORMANCE

In Specification (5) of Table 4, I begin to examine the prior process experience variables of interest by including a measure of the percentage of the team that was hired prior to the announcement of the restocking process change. As previously mentioned, the announcement of the restocking change happened about a month and a half before the actual change. The measure of these Old Guard employees captures both those employees that were in the restocking department previously and those that transferred in after this announcement but before the rollout of the new process. Therefore, non-Old Guard employees include all employees that started working at the store between one and a half months before the restocking change and the focal implementation week. From this specification, it is clear that there is a large penalty associated with higher percentages of Old Guard employees. The coefficient of 61.3 translates to an increase of about 12 missed scans for a one standard deviation increase in the percentage of Old Guard employees. These results corroborate Hypothesis 2 that teams of employees with higher levels of a common history will perform worse immediately after a process change.

It may be the case that the department veterans with direct experience with the old restocking process differ from the transfer veterans with indirect experience with the process. I break out these two types in Specification (6). While the *% Transfer Veteran* measure has a coefficient of 92.973, the *% Dept Veteran* effect is only 54.750. However, given the large standard errors on both of these coefficients, they are statistically indistinguishable from one another.[20] The large standard errors on these coefficients could reflect the relative infrequency of transfer veterans in

[20] Using a rough calculation of the overlap of these coefficients, I find the following 92.793 – 2*34.065 = 24.663 which is less than the mean estimate of the other coefficient (54.750).

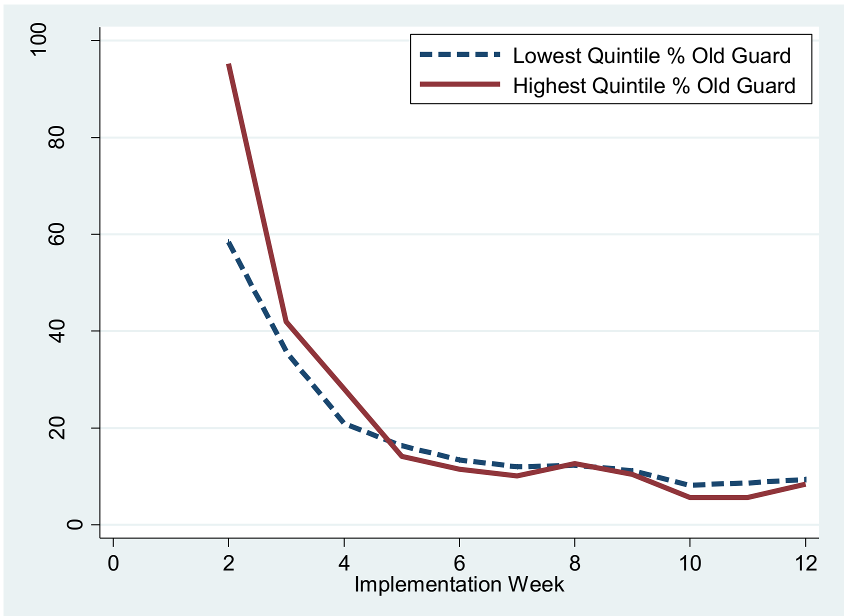
the average store. Because of the indistinguishable difference, I keep the aggregated *% Old Guard* experience measure of these employees going forward.

V.3 PRIOR EXPERIENCE and RATE OF LEARNING

From these specifications, it is clear that teams with higher percentages of employees that have experience with the old restocking process do worse overall, but what about the rate at which they learn? The raw data would suggest that learning rates differ for teams with different exposure to the old restocking practice. Figure 3 plots the average missed scans of those teams that are in the lowest quintile of Old Guard employee representation against the highest quintile. The solid line indicates those teams with the highest Old Guard representation. These teams start off with higher missed scans but improve more rapidly such that these teams are indistinguishable from the other teams in later weeks.[21]

[21] Figure A1 in the Appendix plots the average *% Old Guard* employees per store over time along with the averages of the highest and lowest quintiles *% Old Guard* teams. The graph reveals that these values of *% Old Guard* are fairly stable over the 12 week period, though slightly declining, which suggests that the results I find regarding differences in the performance of these team types cannot be attributed to downward fluctuations in the *% Old Guard* in later weeks.

FIGURE 3
Highest vs Lowest Quintile Teams by % of Old Guard Employees



In order to corroborate this graphical result, I take a slightly different approach to modelling the learning rate. Recall from Figure 2 that the stores improve dramatically at first but then improve much more slowly. Therefore, I fit a piecewise function with one break (two segments) to the data. Using a non-linear least squares model, I find that errors for this model are minimized when the break in the spline occurs after week 4. The resulting specification includes one week variable (*wk*) that is a count of the specific week of implementation and one week variable (*wks5-12*) which takes values of 0 for weeks 2 through 4 but then takes a value of 1 for week 5 (implementation week 5 minus the nonlinear break at week 4), 2 for week 6, and so on. By consolidating the 10 week variables into 2 variables, I am able to interact them with the *% Old Guard* variable to measure a difference in learning rate.

The results examining the learning rate begin in Table 5. In Specification (1), I find that the initial learning rate is quite high, an average improvement of 21.2 scans per week. However, starting in week 5, the improvement slows to only 1.9 scans per week. Specification (3) shows that teams with higher raw average tenures perform worse at first but learn faster than other teams, as reflected by the positive and significant coefficient on the *Emp.Experience* variable and the negative and significant coefficient on the interaction term *Emp.Exper*wk*. These results hold when examining the impact of differences in the concentration of tenure in Specifications (4) and (5).

TABLE 5
Effect of Raw Employee Experience on Learning Rate

	(1) missedscan	(2) missedscan	(3) missedscan	(4) missedscan	(5) missedscan
wk	-22.413***	-21.870***	-18.778***	-18.603***	-21.974***
	(0.795)	(0.805)	(1.870)	(1.879)	(5.986)
wks 5-12 (5=1…)	20.014***	19.602***	16.719***	16.625***	21.040***
	(0.873)	(0.877)	(2.063)	(2.066)	(6.638)
Weekly Sales (000s)	0.015***	0.015***	0.017***	0.017***	0.018***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Avg Units (per trans)	-1.568	-1.770*	-2.083**	-2.110**	-2.103**
	(1.052)	(1.051)	(1.056)	(1.057)	(1.057)
Emp.Experience (yr)		5.086***	8.079***	6.392**	6.234**
		(1.315)	(2.039)	(2.726)	(2.745)
Supervisor Experience (yr)		-0.301	-0.452	-0.360	-0.366
		(1.209)	(1.209)	(1.213)	(1.213)
Emp.Exper*wk			-0.912*	-0.962*	-0.865*
			(0.495)	(0.498)	(0.519)
Emp.Exper*wks 5-12			0.811	0.843	0.710
			(0.548)	(0.549)	(0.575)
Experience Shannon				2.400	0.317
				(2.574)	(4.701)
Experience Shannon*wk					0.653
					(1.107)
Experience Shannon*wks 5-12					-0.859
					(1.228)
Constant	110.982***	95.473***	86.922***	80.814***	90.910***
	(5.713)	(11.745)	(12.682)	(14.275)	(24.110)
Fixed Effect:	Store	Store	Store	Store	Store
Obs.	2872	2872	2872	2872	2872
R-Squared	0.515	0.518	0.519	0.520	0.520

Notes: Dependent variable is the missed scan errors. All columns estimated by using a fixed effect model at the store level. The *wk* variable takes values equal to the implementation week (2, 3, 4, ...12) whereas the *wks 5-12* variable takes values of 0 for implemenation weeks 2-4 then takes values of the implementation week minus 4 (or a value of 1 for week 5, 2 for week 6, etc). *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

In Table 6, I examine whether teams with higher percentages of Old Guard employees learn faster or slower. I find that these teams perform worse at first but also improve much more quickly in the initial weeks. The average improvement rate in the initial weeks is only about 8.9 missed scans generally, but that rate increases dramatically, up to an improvement of 30 missed scans per week, in teams that are composed of completely Old Guard veterans (Specification (3)). The accelerated learning rate associated with the Old Guard employees attenuates after the fourth week and becomes indistinguishable from zero [-21.159+20.836 for each increased week]. These results seem to support Hypothesis 1 that teams of employees with higher levels of common history will learn more quickly.[22] This increased learning rate could be consistent with teams where team members are more exposed to old way of restocking share a reference, and perhaps an ability and language to communicate, that allows for greater improvement each week. This result is also checked against the separated veteran variables to ensure that the results hold for both the combined *%Old Guard* variable interaction and when using separate *department veteran* and *transfer veteran* variables.

[22] Since *Weekly Sales Volume* is somewhat positively correlated with the *%Old Guard* (see Table 2), I separately tested the Old Guard learning rate when including two weekly sales volume and implementation week interactions term. I found the results to be robust and nearly quantitatively identical; therefore, I do not include the regression here.

TABLE 6

Effect of Prior Experience with Old Way on Learning Rate

	(1)	(2)	(3)
	missedscan	missedscan	missedscan
wk	-21.230***	-20.989***	-8.869***
	(0.821)	(0.821)	(2.858)
wks 5-12 (5=1…)	19.375***	19.106***	6.909**
	(0.877)	(0.877)	(3.137)
Weekly Sales (000s)	0.014***	0.016***	0.018***
	(0.005)	(0.005)	(0.005)
Avg Units (per trans)	-1.898*	-2.007*	-2.520**
	(1.049)	(1.046)	(1.052)
Emp.Experience (yr)	-3.732	-16.870***	-3.325
	(2.692)	(4.156)	(2.683)
Supervisor Experience (yr)	-0.030	0.015	-0.138
	(1.208)	(1.204)	(1.202)
% Old Guard	64.775***	35.028*	135.570***
	(17.271)	(18.656)	(22.712)
% Old Guard*Emp.Exper		15.024***	
		(3.629)	
% Old Guard*wk			-21.159***
			(4.752)
% Old Guard*wks 5-12			20.836***
			(5.256)
Constant	85.342***	113.500***	44.891***
	(12.023)	(13.781)	(14.659)
Fixed Effect:	Store	Store	Store
Obs.	2872	2872	2872
R-Squared	0.521	0.524	0.526

Notes: Dependent variable is the missed scan errors. All columns estimated by using a fixed effect model at the store level. The *wk* variable takes values equal to the implementation week (2, 3, 4, ...12) whereas the *wks 5-12* variable takes values of 0 for implemenation weeks 2-4 then takes values of the implementation week minus 4 (or a value of 1 for week 5, 2 for week 6, etc). *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

V.4 DYNAMICS OF EMPLOYMENT

Because the setting for this analysis is an industry where turnover and seasonal employment are known to be high, I consider whether the main results are being driven by stores where employee entry and exit are more prevalent. Prior work has shown that disruptions to team stability are detrimental to performance (Huckman, Staats, and Upton 2009; Reagans, Argote, and Brooks 2005). Therefore, one might be concerned that the results might not be reflective of attributes of teams with higher percentages of Old Guard employees but instead might be driven by those stores that have more turnover during the time period for the data. In order to test this alternative condition, I include additional measures of team stability in Table 7.

Broadly, I find that differences in team size and team stability do matter for performance though the main results remain significant. In Specification (1), I find that the coefficient on the *Employee Count* variable is negative and significant, meaning that larger teams miss fewer scans than smaller teams when controlling for the amount of sales volume in a given week. In Specification (2), I find that each additional new employee in a given store-week is associated with about 1 extra missed scan per week. This result seems logical because new employees inherently know less about the store than more experienced employees and disrupt the existing coordination and communication patterns of the team. The introduction of a new employee, or several new employees, in a given week could be expected to disrupt the new process learning since existing employees must spend time assisting the new employee(s).

TABLE 7

Dynamic Employment Conditions Matter

	(1) missedscan	(2) missedscan	(3) missedscan	(4) missedscan	(5) missedscan	(6) missedscan	(7) missedscan
wk3	-30.026***	-29.967***	-29.948***	-29.878***	-29.891***	-29.989***	-30.027***
	(1.870)	(1.869)	(1.869)	(1.867)	(1.868)	(1.870)	(1.868)
wk4	-41.204***	-40.996***	-41.203***	-40.982***	-40.990***	-41.005***	-41.080***
	(1.931)	(1.932)	(1.930)	(1.930)	(1.930)	(1.932)	(1.931)
wk5	-49.948***	-49.614***	-49.973***	-49.620***	-49.583***	-49.652***	-49.774***
	(2.108)	(2.112)	(2.107)	(2.110)	(2.111)	(2.114)	(2.112)
wk6	-52.960***	-52.676***	-52.946***	-52.643***	-52.597***	-52.704***	-52.783***
	(2.094)	(2.096)	(2.093)	(2.095)	(2.096)	(2.098)	(2.096)
wk7	-52.987***	-52.482***	-52.969***	-52.430***	-52.387***	-52.490***	-52.510***
	(2.106)	(2.116)	(2.105)	(2.115)	(2.116)	(2.117)	(2.115)
wk8	-53.790***	-53.334***	-53.707***	-53.216***	-53.166***	-53.361***	-53.391***
	(2.245)	(2.252)	(2.244)	(2.251)	(2.252)	(2.253)	(2.251)
wk9	-56.840***	-56.325***	-56.751***	-56.196***	-56.119***	-56.364***	-56.382***
	(2.434)	(2.443)	(2.433)	(2.441)	(2.444)	(2.445)	(2.441)
wk10	-58.888***	-58.396***	-58.804***	-58.275***	-58.232***	-58.429***	-58.408***
	(2.561)	(2.568)	(2.560)	(2.567)	(2.567)	(2.570)	(2.567)
wk11	-58.839***	-58.070***	-58.754***	-57.929***	-57.839***	-58.140***	-58.131***
	(2.694)	(2.713)	(2.692)	(2.711)	(2.714)	(2.718)	(2.711)
wk12	-61.285***	-60.242***	-61.336***	-60.229***	-60.155***	-60.273***	-60.319***
	(3.244)	(3.273)	(3.241)	(3.270)	(3.272)	(3.274)	(3.271)
Weekly Sales (000s)	0.017***	0.017***	0.017***	0.017***	0.018***	0.017***	0.017***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Avg Units (per trans)	-2.350**	-2.371**	-2.332**	-2.353**	-2.367**	-2.369**	-2.352**
	(1.076)	(1.075)	(1.075)	(1.074)	(1.074)	(1.075)	(1.074)
Emp.Experience (yr)	-22.242***	-22.060***	-22.113***	-21.910***	-21.725***	-22.131***	-22.347***
	(5.137)	(5.134)	(5.134)	(5.130)	(5.137)	(5.137)	(5.133)
Supervisor Experience (yr)	0.215	0.234	0.194	0.213	0.207	0.257	0.227
	(1.194)	(1.193)	(1.193)	(1.192)	(1.192)	(1.194)	(1.192)
% Old Guard	14.882	18.885	10.212	14.142	16.904	19.263	14.240
	(20.681)	(20.738)	(20.777)	(20.820)	(21.173)	(20.760)	(20.847)
% Old Guard*Emp.Experience	18.350***	18.044***	18.371***	18.047***	17.822***	18.074***	18.315***
	(4.046)	(4.045)	(4.043)	(4.041)	(4.054)	(4.046)	(4.044)
Employee Count	-0.906*	-1.009**	-1.114**	-1.239**	-1.183**	-0.998**	-1.201**
	(0.504)	(0.506)	(0.513)	(0.515)	(0.521)	(0.507)	(0.514)
New Employee		0.914**		0.973**	0.980**	0.914**	0.944**
		(0.400)		(0.400)	(0.401)	(0.400)	(0.400)
Employee Exit			-1.143**	-1.224**			
			(0.526)	(0.527)			
Exit (emp<1yr)					-1.387**		
					(0.573)		
Exit (emp>1yr)					-0.256		
					(1.445)		
Exit (Conduct Violation)						1.008	
						(2.366)	
Exit (Resigned)							-1.505**
							(0.730)
Constant	116.197***	114.924***	122.751***	121.861***	119.134***	114.419***	121.921***
	(25.041)	(25.026)	(25.204)	(25.183)	(25.469)	(25.059)	(25.240)
Fixed Effect:	Store	Store	Store	Store	Store	Store	Store
Obs.	2872	2872	2872	2872	2872	2872	2872
R-Squared	0.536	0.537	0.537	0.538	0.538	0.537	0.538

Notes: Dependent variable is the missed scan errors. All columns estimated by using a fixed effect model at the store level. The *wk#* variables take a value of 1 when it is the focal week and 0 otherwise (e.g. *wk3* = 1 for week 3 and 0 otherwise). The omitted week is week 2. *New Employee* and *Employee Exit* variables are count variables taking positive values when a new employee starts or an employee exits, respectively, and taking a value of 0 otherwise. *Exit (Emp<1yr)* takes positive values if exiting employee(s) have worked for the company for less than one year, and *Exit(Emp>1yr)* tak positive values for those exiting employees with greater than one year of experience. *Exit (Conduct Violation)* takes a value other than 0 when the exiting employee(s) leave as a result of a conduct violation (count variable of number of employees exiting each week). Similarly, *Exit (Resigned)* is a count variable of the number of employees that resign in a given store-week. *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

In the rest of Table 7, I examine the impact of employee exits in detail. While employee entry means that new employees will need to be taught the language of the department, employee exits lower manpower but do not cause a disruption to the language used on the restocking team. From Specifications (3) and (4), it is clear that employee exits do not have the same detrimental effect as new employee entry. Rather, exits seem to improve performance. This result seems surprising at first but becomes perhaps less so upon further examination. There are many reasons for employee exit – some voluntary and some involuntary. In Specification (5), I first examine which employees are leaving and find that the main improvement in missed scans is attributed to exits for those employees that have under 1 year of tenure at the firm. One explanation for this improvement could be that employees who are ill-suited for work on these teams are either leaving or being fired. Another explanation for this improvement could be that employees who are less well-versed in the communication patterns of the larger team are the ones leaving, meaning that the overall communication on the team is strengthened upon employee exit. I further decompose the employee exit characteristic in Specifications (6) and (7) and note that there is a positive, though insignificant increase in missed scans when employees are fired due to a conduct violation (Specification (6)) as opposed to a negative and significant decrease in missed scans when employees give notice and resign (Specification (7)). In the case of employee exit due to resignation, each additional employee that resigns in a given week reduces the errors by about 1.5 scans. This result provides some support for the idea that there is variation in the extent to which employee exits can be disruptive or beneficial. The improvement in performance upon employee exit by those employees less likely to be well-versed in the language of the restocking team provides additional evidence that shared language and increased efficiencies in coordination by Old Guard employees may drive the observed differences in initial performance and learning of different restocking teams.

In Table 8, I reexamine the impact of team size and disruptions to team stability as related to differences in rate of learning. First I examine whether teams of different sizes appear to learn differently. In contrast to Specification (1) of Table 7, I do not find a significant relationship between *Employee Count* and performance when using the nonlinear spline model in Table 8 Specification (1). Further teams of differing sizes do not appear to learn the new process any faster or slower than one another (Specification (2)).

TABLE 8

Dynamic Employment Effects Vary

	(1) missedscan	(2) missedscan	(3) missedscan	(4) missedscan	(5) missedscan	(6) missedscan
wk	-8.811***	-7.135*	-8.598***	-5.982*	-8.762***	-5.883*
	(2.864)	(4.163)	(2.859)	(3.062)	(2.934)	(3.059)
wks 5-12 (5=1…)	6.869**	4.732	6.784**	4.359	6.702**	4.255
	(3.139)	(4.569)	(3.135)	(3.342)	(3.217)	(3.340)
Weekly Sales (000s)	0.018***	0.017***	0.018***	0.018***	0.018***	0.018***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Avg Units (per trans)	-2.535**	-2.472**	-2.561**	-2.689**	-2.568**	-2.681**
	(1.053)	(1.073)	(1.052)	(1.051)	(1.053)	(1.050)
Emp.Experience (yr)	-3.784	-3.925	-3.784	-3.441	-3.652	-3.315
	(2.992)	(3.027)	(2.988)	(2.987)	(2.991)	(2.985)
Supervisor Experience (yr)	-0.115	-0.132	-0.115	-0.237	-0.122	-0.254
	(1.204)	(1.210)	(1.202)	(1.201)	(1.204)	(1.200)
% Old Guard	134.269***	133.887***	132.879***	144.315***	131.023***	140.309***
	(23.023)	(23.042)	(23.076)	(23.169)	(23.095)	(23.214)
% Old Guard*wk	-21.277***	-21.262***	-21.475***	-23.972***	-21.577***	-24.186***
	(4.765)	(4.768)	(4.758)	(4.845)	(4.772)	(4.841)
% Old Guard*wks 5-12	20.918***	20.866***	21.033***	23.413***	21.320***	23.645***
	(5.263)	(5.266)	(5.255)	(5.341)	(5.271)	(5.338)
Employee Count	-0.159	0.047	-0.548	-0.148	-0.345	-0.389
	(0.458)	(0.740)	(0.472)	(0.462)	(0.470)	(0.473)
Employee Count*wk		-0.087				
		(0.156)				
Employee Count*wks 5-12		0.113				
		(0.172)				
New Employee			0.962**	7.183**		7.134**
			(0.402)	(2.981)		(2.978)
Employee Exit			-1.371***		-3.588	-1.235**
			(0.531)		(4.974)	(0.532)
New Employee*wk				-1.407*		-1.396*
				(0.837)		(0.836)
New Employee*wks 5-12				1.011		1.026
				(0.934)		(0.933)
Exit Employee*wk					0.389	
					(1.376)	
Exit Employee*wks 5-12					-0.166	
					(1.494)	
Constant	50.149**	47.009**	57.637***	39.058*	56.749***	46.084**
	(21.089)	(23.837)	(21.247)	(21.381)	(21.332)	(21.576)
Fixed Effect:	Store	Store	Store	Store	Store	Store
Obs.	2872	2872	2872	2872	2872	2872
R-Squared	0.526	0.526	0.528	0.529	0.527	0.530

Notes: Dependent variable is the missed scan errors. All columns estimated by using a fixed effect model at the store level. The *wk* variable takes values equal to the implementation week (2, 3, 4, ...12) whereas the *wks 5-12* variable takes values of 0 for implemenation weeks 2-4 then takes values of the implementation week minus 4 (or a value of 1 for week 5, 2 for week 6, etc). *New Employee* and *Employee Exit* variables are count variables taking positive values when a new employee starts or an employee exits, respectively, and taking a value of 0 otherwise. *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

Because disruptions to team stability have been shown to differentially affect performance of teams based on the level of adherence to the prescribed process (Ton and Huckman 2008), I also examine the impact of disruptions in different weeks of the implementation in Table 8. Specifically, Ton and Huckman (2008) find that teams who deviate from the prearranged process have worse performance after employee turnover than those teams that follow the process more closely. This finding would suggest that in the early weeks when teams are still new to the process, disruptions might be more detrimental than in later weeks when the process has been executed more times. In Specification (2), I do find that new employee entry is most detrimental in the first weeks after the implementation, but I do not find evidence of a similar effect for exiting employees in Specification (3). Most importantly, the previous results regarding the performance and learning rate of teams of Old Guard employees remains consistent and significant throughout this analysis of dynamic employment in retail stores.

VI. DISCUSSION

In this section, I address the basic concern regarding the mechanism that is driving the results and perform some robustness checks on the analysis. While I have argued above that the result may be due to both the existence of basic competency traps experienced by teams that are otherwise prone to efficiencies in learning new tasks, there are other potential mechanisms which I will address and rule out in turn. Also, I provide a set of robustness checks regarding the specific construction of this dataset to rule out any remaining concerns about potential biases.

VI.1 WHY DOES PRIOR EXPERIENCE WITH AN OLD PRACTICE MATTER?

Although competency traps hinder performance in firms, it has been extremely hard to identify instances where competency traps, rather than some other mechanism, are at work. I have shown that teams with increased experience with the old way of restocking suffer worse performance at the time of transition to the new practice, which I suggest may be attributed to competency traps, but learn to improve more quickly, which I suggest may be attributed to increased efficiencies in communication and coordination among teams where individuals know how to work with one another.

In speaking with the manager of a single restocking team, he specifically addressed both of these aspects when discussing his team's response to the new restocking process. First, as mentioned previously with regard to how disruptive the restocking process was for existing employees, the manager suggested that this process is not all that complicated but that employees do "come in expecting to do the same thing every night." Disruptions to existing employees' tasks are necessarily costly because employees operate in predictable patterns of workflow each night. Second, when asked what he sees as the number one reason for the improvement in the process over time, he said that "A lot of people [referring to the employees that come from the headquarters to evaluate the new process] don't realize that this process takes a lot of communication. People have to understand each other and talk or it doesn't work as well."

When asked to explain the communication that is needed and what exemplifies good communication restocking managers described the communication necessary between employees doing different phases of the restocking process. Specifically one cited example was explained to me as follows - the packout team unpacks boxes and loads them onto shelves, leaving the remaining product that does not fit onto the immediate shelving for the packdown team to place above the shelves. The process becomes much smoother if the packout team communicates the locations in which extra products need to be placed above the shelves and where there may be missing product on the shelves that may need to be pulled down from the storage above for the trailing packdown team. There are several of these types of pass-offs between team members throughout the process, and better communication between members executing these pass-offs helps reduce the overall errors in the process.

The managerial response regarding the way the process works lends some evidence that the mechanism for improvement may be better communication and coordination. However, because I cannot observe communication directly in this setting, I quickly consider alternative mechanisms in turn to rule out the possibility of a different explanation for my results. Because of the results, I am only concerned with mechanisms that might be associated with high proportion Old Guard teams performing both worse at first and improving more quickly over time. In this setting, there is no performance pay or other way in which teams differ along the dimension of extrinsic motivation. The main considerations for alternative mechanisms are whether different types of employees are more or less competent and more or less intrinsically motivated. First, if employees differ in their ability to perform the new process because they differ in level of competency, I would expect the teams with higher percentages of Old Guard employees to perform worse across all periods, which is inconsistent with the finding. Second, if employees differ in their intrinsic motivation along any dimension, I would expect that teams with higher percentages of Old Guard employees would learn more slowly than their counterpart teams, which is also inconsistent with the findings.[23]

I am unaware of alternative mechanisms that may explain the main results. Although the *% Old Guard* variable does not measure communication and coordination directly, ruling out alternative mechanisms that may cause differences in team behavior provides additional evidence in support

[23] Teams may differ in intrinsic motivation in three key ways. Employees on these teams may differ in their investments in firm-specific knowledge, in their ability to shirk at their assigned tasks, and in the extent to which they feel "wronged" by this process change.

of my hypotheses. In sum, I reject the notion that teams suffer from differences in level of opportunism (Holmstrom 1982) in favor of an possible explanation that these teams suffer from informational inefficiencies related to coordination (Marshak and Radner 1972).

VI.2 ROBUSTNESS

In the robustness analysis I consider alternative restrictions for outlier exclusion in the data. For the main analysis, I excluded observations where the missed scan performance metric was greater than 400. I chose a cutoff of 400 scans because those falling above this cutoff were significantly higher than other stores' rates and could have experienced implementation issues of a different nature than other stores. This cutoff only excluded 3 observations, all from the initial week of implementation, but it might be of some concern to choose 400 scans as a cutoff versus including all the observations or imposing a stricter criterion for inclusion in the analysis.

Therefore, in Table 9, I replicate the results from Table 6 alternatively including all the data and imposing a cutoff of no more than 200 missed scans. I find that the results remain significant and qualitatively the same across both criteria but the magnitudes of the coefficients change. Quantitatively, the coefficients on *%Old Guard*wk* and *%Old Guard*wks5-12* are not significantly different from the coefficients found in the main result; however, they are significantly different from one another.[24] From the increases in the coefficients on *% Old Guard*, *%Old Guard*wk*, and *%Old Guard*wks5-12* in Specifications (2) and (3), it is clear that the excluded observations in the main analysis were from stores that had high percentages of Old Guard employees. Therefore, if anything, my main results are a more conservative estimate of

[24] The significance of the difference is calculated by the Z-test as follows: $= \frac{\beta_1 - \beta_2}{\sqrt{SE(\beta_1)^2 + SE(\beta_2)^2}}$.

the effect of experience with the old way on performance and learning of the new way of doing things.

TABLE 9
Alternative Outlier Removal

	All Data			Missed Scans <200		
	(1)	(2)	(3)	(4)	(5)	(6)
	missedscan	missedscan	missedscan	missedscan	missedscan	missedscan
wk	-24.844***	-24.604***	-7.045*	-19.060***	-18.830***	-11.612***
	(1.065)	(1.066)	(3.721)	(0.724)	(0.723)	(2.512)
wks 5-12 (5=1…)	23.438***	23.169***	5.139	17.157***	16.900***	9.801***
	(1.136)	(1.138)	(4.084)	(0.773)	(0.772)	(2.756)
Weekly Sales (000s)	0.006	0.008	0.011*	0.013***	0.015***	0.016***
	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)
Avg Units (per trans)	-0.330	-0.439	-1.141	-1.437	-1.541*	-1.849**
	(1.367)	(1.365)	(1.371)	(0.919)	(0.915)	(0.924)
Emp.Experience (yr)	-7.480**	-20.618***	-6.783*	-2.491	-15.193***	-2.291
	(3.506)	(5.423)	(3.494)	(2.355)	(3.632)	(2.352)
Supervisor Experience (yr)	-0.069	-0.023	-0.237	0.150	0.194	0.089
	(1.575)	(1.572)	(1.567)	(1.056)	(1.052)	(1.054)
% Old Guard	79.446***	49.674**	179.527***	55.547***	26.832	99.196***
	(22.499)	(24.343)	(29.523)	(15.115)	(16.309)	(20.000)
% Old Guard*Emp.Exper		15.027***			14.522***	
		(4.738)			(3.170)	
% Old Guard*wk			-30.521***			-12.754***
			(6.180)			(4.187)
% Old Guard*wks 5-12			30.803***			12.226***
			(6.838)			(4.629)
Constant	98.686***	126.850***	41.267**	74.298***	101.505***	49.508***
	(15.675)	(17.992)	(19.101)	(10.529)	(12.053)	(12.878)
Fixed Effect:	Store	Store	Store	Store	Store	Store
Obs.	2875	2875	2875	2866	2866	2866
R-Squared	0.432	0.435	0.439	0.558	0.562	0.561

Notes: Dependent variable is the missed scan errors. All columns estimated by using a fixed effect model at the store level. Specifications (1) - (3) include all data available whereas Specifications (4)-(6) include only those store-weeks where the missed scan errors were less than 200. The *wk* variable takes values equal to the implementation week (2, 3, 4, ...12) whereas the *wks 5-12* variable takes values of 0 for implemenation weeks 2-4 then takes values of the implementation week minus 4 (or a value of 1 for week 5, 2 for week 6, etc). *New Employee* and *Employee Exit* variables are count variables taking positive values when a new employee starts or an employee exits, respectively, and taking a value of 0 otherwise. *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

VII. CONCLUSION

In the context of a retail setting, I examine the impact of prior worker experience with a superseded, but related, practice on the performance and learning of a new practice. This work extends the current learning literature by considering how a firm's history affects the availability of current human capital assets and its future ability to implement new practices. In particular, I have found support that teams that have higher proportions of employees with a shared history executing the old practice will perform worse at the outset of a new practice but will improve more rapidly with repetition of the new practice. These teams initially suffer from a competency trap stemming from their shared history with the old practice but achieve greater efficiencies in communication and coordination that accelerates learning with the new practice. The data are consistent with this mechanism, and informal interviews with store employees and managers indicated that this aspect of coordination and shared understanding between team members is a crucial aspect in team success. However, I also evaluate other mechanisms that may be present in heterogeneous worker teams in order to rule out alternatives that would produce the same results in this setting. While ruling out alternatives is certainly not complete evidence of the validity of my proposed mechanism, it does alleviate particular concerns. In providing evidence for the existence of competency traps in a process change as seemingly basic as restocking product on shelves, I suggest that differences in prior experiences of a firm's human capital are an important aspect of why we see performance differences among seemingly similar firms.

Lastly, I also suggest an alternative method by which learning studies can capture learning empirically. The exponential and power function empirical specifications common to organizational learning studies assume a constant learning rate that is quite restrictive. By assuming a more non-parametric form, scholars can allow for a variable learning rate over time. In this study specifically, it is clear that a constant learning rate would have mis-specified the rate at which stores improve performance. Therefore, a re-evaluation of the constant learning rate in these types of studies seems warranted going forward.

While the setting for this study was chosen purposefully and the results have clear generalizability to other multi-unit firms, applicability in other settings and to other types of practice changes requires more research. In particular, this change in restocking was fairly incremental and the purpose of the task and team remained the same. Future research might examine changes where the practice change is discontinuous and / or occurs in a setting where the employee roles involve more knowledge work. Also, this change was endogenously decided with a defined implementation schedule, but many incremental changes in firms occur as a result of some exogenous change outside the firm's control. When changes occur as a result of factors outside the firm, the response of teams should be explored in greater detail. While this work provides some expected results of these future studies, more research is needed to understand how these conclusions extend to different settings.

BIBLIOGRAPHY

Agarwal, Rajshree, and Constance E. Helfat. 2009. "Strategic Renewal of Organizations." *Organization Science* 20 (2): 281–293.

Allen, Thomas J, and Stephen I Cohen. 1969. "Information Flow in Research and Development Laboratories." *Administrative Science Quarterly* 14 (1): 12–19.

Argote, Linda. 2013. *Organizational Learning: Creating, Retaining and Transferring Knowledge*. Second Edi. New York: Springer.

Argote, Linda, S. L. Beckman, and Dennis Epple. 1990. "The Persistence and Transfer of Learning in Industrial Settings." *Management Science* 36 (2): 140–154.

Argote, Linda, and Dennis Epple. 1990. "Learning Curves in Manufacturing." *Science* 247 (4945): 920–947.

Argote, Linda, and Paul Ingram. 2000. "Knowledge Transfer: A Basis for Competitive Advantage in Firms." *Organizational Behavior and Human Decision Processes* 82 (1): 150–169.

Arrow, Kenneth. 1974. *The Limits of Organization*. New York: Norton.

Bartelsman, Eric J, and Mark Doms. 2000. "Understanding Productivity: Lessons from Longitudinal Microdata." *Journal of Economic Literature* 38 (3): 569–594.

Baum, JAC, and Paul Ingram. 1998. "Survival-Enhancing Learning in the Manhattan Hotel Industry, 1898–1980." *Management Science* 44 (7): 996–1016.

Bayus, B. L., and R. Agarwal. 2007. "The Role of Pre-Entry Experience, Entry Timing, and Product Technology Strategies in Explaining Firm Survival." *Management Science* 53 (12): 1887–1902.

Benkard, C. Lanier. 2000. "Learning and Forgetting: The Dynamics of Aircraft Production." *American Economic Review* 90 (4): 1034–1054.

Chatterji, Aaron K. 2009. "Spawned with a Silver Spoon? Entrepreneurial Performance and Innovation in the Medical Device Industry." *Strategic Management Journal* 30 (2): 185–206.

Chatterji, Aaron K., Rui J.P. de Figueiredo Jr, and Evan Rawley. 2014. "Learning on the Job? Entrepreneurial Spawning in the Asset Management Industry." *Columbia Business School Research Paper No. 14-64*: 1–43.

Chen, Pao-lien, Charles Williams, and Rajshree Agarwal. 2012. "Growing Pains: Pre-Entry Experience and the Challenge of Transition to Incumbency." *Strategic Management Journal* 33: 252–276.

Cohen, WM, and DA Levinthal. 1990. "Absorptive Capacity: A New Perspective on Learning and Innovation." *Administrative Science Quarterly* 35 (1): 128–152.

Darr, Eric D., Linda Argote, and Dennis Epple. 1995. "The Acquisition, Transfer, and Depreciation of Knowledge in Service Organizations: Productivity in Franchises." *Management Science* 41 (11): 1750–1762.

DeHoratius, N., A. J. Mersereau, and L. Schrage. 2008. "Retail Inventory Management When Records Are Inaccurate." *Manufacturing & Service Operations Management* 10 (2): 257–277.

DeHoratius, N., and A. Raman. 2008. "Inventory Record Inaccuracy: An Empirical Analysis." *Management Science* 54 (4): 627–641.

Denrell, Jerker, and JG March. 2001. "Adaptation as Information Restriction: The Hot Stove Effect." *Organization Science* 12 (5): 523–538.

Dutton, JM, and Annie Thomas. 1984. "Treating Progress Functions as a Managerial Opportunity." *Academy of Management Review* 9 (2): 235–247.

Edmondson, Amy C, Ann B Winslow, Richard M J Bohmer, and Gary P. Pisano. 2003. "Learning How and Learning What: Effects of Tacit and Codified Knowlenge on Performance Improvement Following Technology Adoption." *Decision Sciences* 34 (2): 197–224.

Epple, Dennis, Linda Argote, and Kenneth Murphy. 1996. "An Empirical Investigation of the Microstructure of Knowledge Acquisition and Transfer through Learning by Doing." *Operations Research* 44 (1): 77–86.

Felin, Teppo, Nicolai J. Foss, Koen H. Heimeriks, and Tammy L. Madsen. 2012. "Microfoundations of Routines and Capabilities: Individuals, Processes, and Structure." *Journal of Management Studies* 49 (8): 1351–1374.

Fleisch, Elgar, and Christian Tellkamp. 2005. "The Impact of Inventory Inaccuracy on Retail Supply Chain Performance : A Simulation Study White Paper." *International Journal of Production Economics* 95: 373–385.

Grant, Robert M. 1996. "Toward a Knoweldge-Based Theory of the Firm." *Strategic Management Journal* 17 (Winter Special Issue): 109–122.

Hansen, MT. 1999. "The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits." *Administrative Science Quarterly* 44 (1): 82–111.

Harrison, David A., and Katherine J. Klein. 2007. "What's the Difference? Diversity Constructs as Separation, Variety, or Disparity in Organizations." *Academy of Management Review* 32 (4): 1199–1228.

Hatch, NW, and DC Mowery. 1998. "Process Innovation and Learning by Doing in Semiconductor Manufacturing." *Management Science* 44 (11-part-1): 1461–1477.

Hedberg, Bo. 1981. "How Organizations Learn and Unlearn." In *Handbook of Organizational*

Design, edited by P.C. Nystrom and W.H. Starbuck, Vol. 1, 3–27. Oxford, UK: Oxford Univ Press.

Helfat, Constance E. 1994. "Evolutionary Trajectories Firm R & D in Petroleum." *Management Science* 40 (12): 1720–1747.

Helfat, Constance E., and MB Lieberman. 2002. "The Birth of Capabilities: Market Entry and the Importance of Pre-History." *Industrial and Corporate Change* 11 (4): 725–760.

Helfat, Constance E., and Margaret A. Peteraf. 2003. "The Dynamic Resource-Based View: Capability Lifecycles." *Strategic Management Journal* 24 (10): 997–1010.

Herriott, Scott R, Daniel Levinthal, and James G March. 1985. "Learning from Experience in Organizations." *The American Economic Review* 75 (2): 298–302.

Hitt, Michael, Leornard Bierman, Katsuhiko Shimizu, and Rahul Kochhar. 2001. "Direct and Moderating Effects of Human Capital in Professional on Strategy and Performance Service Firms : A Resource-Based Perspective." *Academy of Management Journal* 44 (1): 13–28.

Holmstrom, Bengt. 1982. "Moral Hazard in Teams." *The Bell Journal of Economics* 11 (2): 74–91.

Huckman, R. S., B. R. Staats, and D. M. Upton. 2009. "Team Familiarity, Role Experience, and Performance: Evidence from Indian Software Services." *Management Science* 55 (1): 85–100.

Kang, Yun, and Stanley B. Gershwin. 2005. "Information Inaccuracy in Inventory Systems: Stock Loss and Stockout." *IIE Transactions* 37 (9): 843–859.

Klein, J I. 1989. "Parenthetic Learning in Organizations: Toward the Unlearning of the Unlearning Model." *Journal of Management Studies* 26 (3): 291–308.

Klepper, Steven, and Kenneth L. Simons. 2000. "Dominance by Birthright: Entry of Prior Radio Producers and Competitive Ramifications in the U.S. Television Receiver Industry." *Strategic Management Journal* 21 (10-11): 997–1016.

Klepper, Steven, and Sally Sleeper. 2005. "Entry by Spinoffs." *Management Science* 51 (8): 1291–1306.

Lapré, Michael, and Nikos Tsikriktsis. 2006. "Organizational Learning Curves for Customer Dissatisfaction: Heterogeneity Across Airlines." *Management Science* 52 (3): 352–366.

Leonard-Barton, Dorothy. 1992. "Core Capabilities and Core Rigidities: A Paradox in Managing New Product Development." *Strategic Management Journal* 13 (1): 111–125.

Levitt, Barbara, and James G. March. 1988. "Organizational Learning." *Annual Review of Sociology* 14 (1): 319–338.

Levitt, SD, JA List, and Chad Syverson. 2013. "Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant." *Journal of Political Economy* 121 (4):

643–681.

Lieberman, MB. 1984. “The Learning Curve and Pricing in the Chemical Processing Industries.” *The RAND Journal of Economics* 15 (2): 213–228.

Lounamaa, P. H., and J. G. March. 1987. “Adaptive Coordination of a Learning Team.” *Management Science* 33 (1): 107–123.

March, James. 1991. “Exploration and Exploitation in Organizational Learning.” *Organization Science* 2 (1): 71–87.

March, James, and Herbert Simon. 1958. *Organizations*. New York: John Wiley & Sons.

Marshak, Jacob, and Roy Radner. 1972. *Economic Theory of Teams*. New Haven: Yale Press.

Moylan, Martin. 2015. “Expect More? Empty Shelves Frustrate Target Customers, Execs.” *MPR News*. September 3. http://www.mprnews.org/story/2015/09/03/empty-target-shelves.

Nelson, Richard, and Sidney Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, MA: Harvard University Press.

Nystrom, Paul C., and William H. Starbuck. 1984. “To Avoid Organizational Crises, Unlearn.” *Organizational Dynamics* 12 (4): 53–65.

Pisano, Gary P. 1994. “Knowledge, Integration, and the Locus of Learning: An Empirical Analysis of Process Development.” *Strategic Management Journal* 15: 85–100.

Pisano, Gary P., Richard M.J. Bohmer, and Amy C. Edmondson. 2001. “Organizational Differences in Rates of Learning: Evidence from the Adoption of Minimally Invasive Cardiac Surgery.” *Management Science* 47 (6): 752–768.

Rapping, L. 1965. “Learning and World War II Production Functions.” *The Review of Economics and Statistics* 47 (1): 81–86.

Reagans, Ray, Linda Argote, and Daria Brooks. 2005. “Individual Experience and Experience Working Together: Predicting Learning Rates from Knowing Who Knows What and Knowing How to Work Together.” *Management Science* 51 (6): 869–881.

Rhee, Mooweon, and Tohyun Kim. 2015. “Great Vessels Take a Long Time to Mature: Early Success Traps and Competences in Exploitation and Exploration.” *Organization Science* 26 (1): 180–197.

Rosenbloom, RS, and CM Christensen. 1994. “Technological Discontinuties, Organizational Capabilities, and Strategic Commitments.” *Industrial and Corporate Change* 3(3): 655–685.

Saporito, Bill. 2013. “The Trouble Lurking on Walmart’s Empty Shelves.” *TIME.com*.

Schilling, Melissa A., Patricia Vidal, Robert E. Ployhart, and Alexandre Marangoni. 2003. “Learning by Doing Something Else: Variation, Relatedness, and the Learning Curve.” *Management Science* 49 (1): 39–56.

Shafer, SM, D Nembhard, and M Uzumeri. 2001. "The Effects of Worker Learning, Forgetting, and Heterogeneity on Assembly Line Productivity." *Management Science* 47 (12): 1639–1653.

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

Szulanski, Gabriel. 1996. "Exploring Internal Stickiness: Impediments to the Transfer of Best Practice within the Firm." *Strategic Management Journal* 17(Special Issue: Knowledge and the Firm): 27–43.

Teece, David J. 2007. "Explicating Dynamic Capabilitites: The Nature and Microfoundations of (Sustainable) Enterprise Performance." *Strategic Management Journal* 28 (13): 1319–1350.

Teece, David J, Gary Pisano, and Amy Shuen. 1997. "Dynamic Capabilities and Strategic Management." *Strategic Management Journal* 18 (7): 509–533.

Thompson, Peter. 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.

Ton, Zeynep, and Robert S. Huckman. 2008. "Managing the Impact of Employee Turnover on Performance: The Role of Process Conformance." *Organization Science* 19 (1): 56–68.

Tushman. 1978. "Technical Communication in R&D Laboratories: The Impact of Project Work Characteristics." *Academy of Management Journal* 21 (4): 624–645.

Tushman, ML, and P Anderson. 1986. "Technological Discontinuities and Organizational Environments." *Administrative Science Quarterly* 31 (3): 439–465.

Weber, RA, and CF Camerer. 2003. "Cultural Conflict and Merger Failure: An Experimental Approach." *Management Science* 49 (4): 400–415.

Wright, Theodore Paul. 1936. "Learning Curve." *Journal of the Aeronautical Sciences* 3 (1): 122–128.

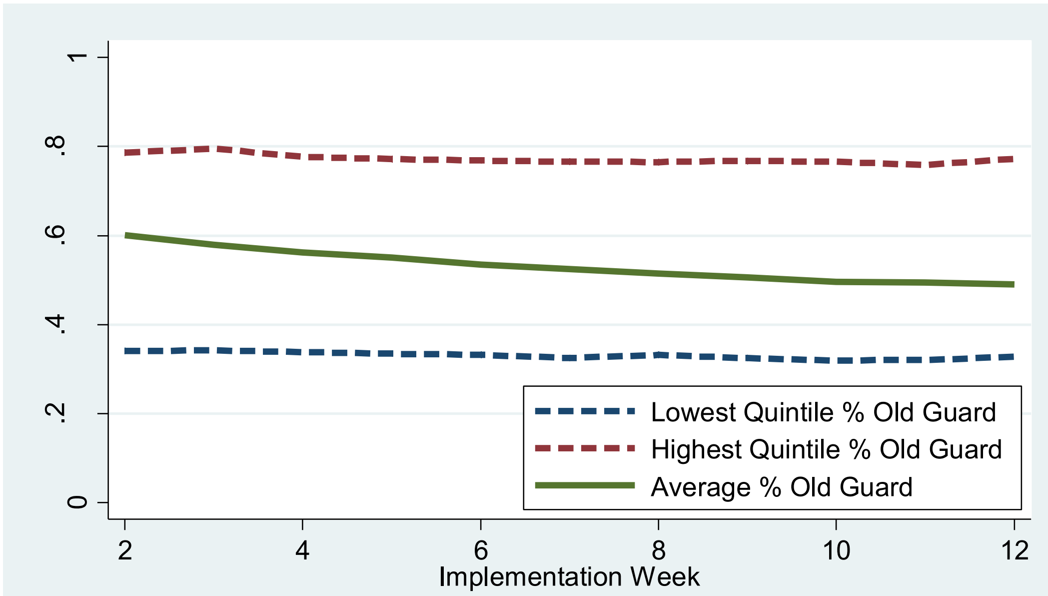
Yelle, Louis E. 1979. "The Learning Curve: Historical Review and Comprehensive Survey." *Decision Sciences* 10 (2): 302–328.

Zenger, T. R., and B. S. Lawrence. 1989. "Organizational Demography: The Differential Effects of Age and Tenure Distributions on Technical Communication." *Academy of Management Journal* 32 (2): 353–376.

APPENDIX

APPENDIX FIGURE 1

Average % Old Guard per Restocking Team



Essay 3 | **How Do Pilots Work?**

Examining Pilot Use in New Practice Transfer

Megan Lawrence

ABSTRACT

This paper examines how the use of pilots influences new practice implementation in a multi-unit firm. Using data from a Fortune 100 retail chain that implemented a new restocking process in its stores, I compare the implementation performance of the same restocking practice in 280 stores. These stores are divided into districts, each with one pilot store that implemented the new process first. By examining how prior performance relates to current performance, I propose and subsequently test the relationship that pilots enable learning from outside experiences in practice transfer and implementation. I find that stores only learn from the pilot store in their own district, even when they are closer to another district's pilot. However, stores' prior performances relative to their pilot stores' prior performances influence the extent to which they learn from their own experiences versus from their pilot stores. Overall, these findings suggest that (1) pilot use for transferring practices does influence performance; (2) learning from other stores is moderated by performance feedback; and (3) contextual differences matter in the context of transferring practices. Lastly, I include initial work providing some evidence that the most likely alternative mechanism in this setting – peer effects – may also be present but that these effects do not eliminate the learning findings.

I. INTRODUCTION

Replication of existing practices within firms is an important value-creating strategy. In fact, one of the main benefits of organizing a business into a retail chain is the ability to produce economies of scale or scope through replication of a standardized set of practices and specifications (Chandler 1977). The ability to access knowledge gained from the chain's experience and to transfer knowledge and practices from one location to another (Kogut and Zander 1992) is considered a key driver explaining why these firm types tend to centralize decision-making (Chang and Harrington 2000). However, the actual method of transfer of knowledge and practices, even among centralized wholly-owned and managed firms, is not clear. Among the several alternatives which include tactics like transferring employees across locations and relying on transfer through geographic proximity, one specific way firms try to transfer practices and to overcome variation in adoption is to use working models or pilot implementations of a new practice. While these pilot implementations are widely used, their efficacy for practice transfer and implementation purposes is not well understood.

The use of pilot implementations is related to the method proposed by Nelson and Winter (1982) of using a working example, or 'template', as a mechanism for practice transfer. While there exists some evidence that template use and local training may increase the efficacy of the adoption of new practices (Jensen and Szulanski 2007; Bloom et al. 2013), the impact of pilot use in new practice implementations and the mechanism(s) by which a pilot or template actually influence(s) its receiving unit is/are not well understood. In this paper, I consider how the use of pilot stores in the rollout of a new organizational practice impacts new practice performance for receiving units. More specifically, I propose that replicating units learn to implement a specific new practice by using both their own experiences and those of the pilot to inform improvement, suggesting that the main purpose and function of the pilot store is to provide an opportunity to learn from another's experience rather than relying solely on one's own experience in practice transfer. This notion that units within a multi-unit firm learn from one another, and perhaps specifically from a focal model unit when transferring and implementing new practices, is consistent with the notion of value creation through replication in these firm types.

Generally, when considering learning within firms, scholars have focused on two categories of learning in the implementation of new practices – learning from one's own experience and learning from another's experience. Learning from one's own experience is often referred to as learning-by-doing (see review by Argote 2013) whereas indirect learning or learning from another's experience is often also called vicarious learning (Levitt and March 1988; Argote and Ingram 2000). These two types of learning are most often considered separately with the focus being on the determinants and moderators of learning speed and/or learning spillovers. Relatively little work has focused on the interaction of these two types (Schwab 2007; Baum and Dahlin 2007). However, if pilots are useful as continuing points of reference in new practice implementation, these two learning types likely occur simultaneously.

In addition to examining multiple forms of learning, I consider how these learning modes are impacted by intra-organizational frictions. In doing so, I suggest that different forms of learning are interdependent rather than sequential, as indicated by Bingham and Davis (2012), occurring simultaneously with implementation teams adjusting how they learn based on their limited information about their own, as compared to their contemporaries', performance. The reliance on one's own experience or other's experience as interdependent and performance-based suggests that adopting new practices in multi-unit firms by learning from pilots still results in a fair amount of intra-firm performance variation.

I study these relationships in the context of a specific and identical new practice implementation occurring simultaneously across many stores within a single firm. I examine how a store's learning of the new practice is impacted by both its own experience and that of its local pilot using internal data from a Fortune 100 retail firm that made a change to the way it restocks its stores. The particular way in which the process is rolled out in firm stores as well as the organizational structure of this firm provide an opportunity to examine both how the stores learn from their own experiences as well as how the stores learn from their pilot store and are perhaps additionally influenced by their peer groupings. Stores are divided into predetermined local districts – groupings of seven to eleven stores – that are used for oversight and management purposes. Taking into consideration this division of units, I am able to examine when and how stores learn from their designated pilot store, which implemented the process a few weeks early, and, to some extent, to examine the impact of performance of peer stores which implemented the practice at the same time.[25] The impact of peer stores on focal store performance is a form of peer effect that may influence a store's motivation or effort to improve performance, and I briefly consider this alternative mechanism; however future work is needed to conclusively separate potential peer effect mechanism from the candidate learning mechanism.

I find that the local district in which a store is located and its local pilot store's performance have an impact on the store performance of the new practice. Stores learn from their own district pilot store but not from more geographically proximate pilot stores. Further, learning from the pilot

[25] Given that there is only one restocking team per store, I use the terms "store" and "team" interchangeably.

store is tempered by the performance feedback a store receives about its own performance as compared to the pilot's performance. Stores value their own performance more heavily and place no significant emphasis on prior pilot store performance when they perform better than the pilot store in the previous week. The extent to which stores value their own prior performance is increased with increasing frictions to the relationship between pilot and nonpilot stores – frictions include differences in the store sales volume and type of customers served as measured by the restocking shift hour overlaps and population density differences in the areas near the two stores.

Lastly, restocking teams are somewhat sensitive to information about how well they perform at the restocking task in the prior week as compared to their district neighbors, but initial evidence suggests that peer effects do not dominate the relationship found between stores and their pilots. Stores that performed worst in their district overall the previous week are much more likely to improve compared to their peers in the following week. Their peers tend to persist in their performance relative to one another week-over-week. However, when examining week-over-week improvement in performance as a function of the specific rank ordering of stores in the district, there is no evidence that performance improves as a function of being in key ranking locations such as best-in-district, worst-in-district, or second-worst-in-district.

Overall, these findings support a notion that acting as a model, or pilot, to promote vicarious learning is one significant function of pilot stores. Further, learning from one's own and from other's experience in practice transfer seems to occur simultaneously and adjust with as little as a week's lag based on the local intra-firm information. Perhaps most importantly, the chosen local pilot store influences other's new practice performance not just by its own ability to perform but also because of its similarity to its local group along non-performance dimensions.

This paper contributes to our understanding of the specific function of pilot implementations in firms as well as generally how firms are able to adopt new practices and build the capabilities that lead to sustained competitive advantages in firms. In exploring the dynamic response of units to own and other experience, I suggest that performance feedback information may dictate the extent to which units look to improve from direct versus indirect experiences with new practices as opposed to the relative exposure that these units have to the new practice. Because experience is a fundamental component in learning (Cyert and March 1963; Huber 1991) and learning is critical in understanding how firms adapt and develop new competencies (Teece 2007; Helfat and Winter 2011), it is important to consider the types, and relationships between types, of experience that firms and teams draw from when implementing new practices. Understanding these differences contributes to an overall understanding of how firms may differ in both practice choice and practice performance over time and, ultimately, why we continue to see persistent performance differences among seemingly similar firms (Gibbons and Henderson 2012).

II. LITERATURE and THEORY

II.1 REPLICATION OF PRACTICES

In the context of multi-unit firms, one of the ways that corporate parents create value is through the learning and knowledge transfer that occur through owning multiple units. This value is accumulated both in selecting new practices to implement and in facilitating adoption of these practices by each unit. Regarding the latter – in Winter and Szulanski (2001), the phrase "replication as strategy" came to mean that replication of the same practice across locations could be considered a way to create value in firms. This replication of the same practice is a way of transferring the valuable knowledge that is embedded in the practice (Nelson and Winter 1982; Prahalad and Hamel 1990; Zollo and Winter 2002) to different contexts and even markets (Helfat and Peteraf 2003; March 1991; Szulanski and Winter 2002).

The study of replication itself has spanned many different contexts and governance models – including subsidiaries in a multinational context (Jensen and Szulanski 2004), franchise organizations (e.g. Darr, Argote, and Epple 1995; Kalnins and Mayer 2004), and business units of corporate parents (Parmigiani and Holloway 2011). In investigating learning and knowledge transfer in multi-unit firms, scholars have often focused on how the governance mode impacts learning and have found that knowledge transfers more easily within than across systems (Darr, Argote, and Epple 1995; Ingram and Baum 1997; Sorenson and Sorensen 2001; Kalnins and Mayer 2004). There has been relatively little examination of how this replication of learnings transfers within a single ownership system or what tools are used in the transfer.

The use of templates is one proposed method of transfer (Nelson and Winter 1982) that seems most closely related to the observed business practice of using pilot implementations prior to the rollout of a new practice across all units. Some limited evidence exists that the template type of transfer increases efficacy of the adoption of practices (Jensen and Szulanski 2007). However, when replicating knowledge and activities across locations; context matters a great deal in the success of replication (Knott 2003; Williams 2007; Gupta, Hoopes, and Knott 2014). Even minute deviations from the existing templates in some instances can cause large performance penalties (Rivkin 2000; Rivkin 2001). Despite suggestive evidence that this type of replication tool is perhaps beneficial under some circumstances, and at the very least is widely used, relatively little empirical research considers the role of pilots in new practice transfer specifically or the potential environments where the efficacy of pilot implementations may be decreased.

II.2 EXPERIENTIAL and VICARIOUS LEARNING

Learning in this context is defined as a regular shift in behavior or knowledge that is influenced by some previous experience (Cyert and March 1963; Levitt and March 1988). The two types of learning most relevant for practice transfer are experiential learning and vicarious learning. Both of these types of learning share the foundation of trying to replicate the successful aspects and avoid the unsuccessful aspects of an implementation (March 2010).

First, there is much evidence that firms can improve at a given task just by doing it. This type of improvement is most often characterized by a learning curve, and estimates of this type of learning have been captured across a wide variety of settings (Argote, Beckman, and Epple 1990; Benkard 2000; Wright 1936; Pisano, Bohmer, and Edmondson 2001; Reagans, Argote, and Brooks 2005; Darr, Argote, and Epple 1995). Generally, studies of this type of learning take performance improvement as a given and focus on attributes of the organization or learning context that impact the rate of improvement. While learning from one's own experience is not a guarantee and scholars have repeatedly asserted the challenge of learning from experience (Levitt and March 1988; Levinthal and March 1993), the general benefit of increased experience on regular task or practice performance is fairly well established.

In addition to learning from one's own experience, firms and teams often learn from the experiences of others. This type of learning, also known as vicarious learning, is indirect in that the firm itself is not completing the practice but rather altering its behaviors in response to the behaviors of or direction from other firms (Huber 1991; Levitt and March 1988; Haunschild and Miner 1997; Bresman 2013; Szulanski 1996; Argote and Ingram 2000). These other firms are often perceived as having some sort of expertise or additional knowledge in an environment where the practice, product, or task has some uncertainty. The value of others' experiences depends on the comparability of the two groups and the situations faced (Greve 1998). In this particular instance where the model pilot team is clearly delineated and given the opportunity to learn the new restocking practice first, the headquarters makes a purposeful effort to create an environment for sharing useful information between the pilot teams and the other district teams.

These two types of learning in firms, experiential and vicarious, have often been considered separately or as occurring in sequence (Schwab 2007; Bingham and Davis 2012). However, learning theories in general have suggested that the patterns of learning adjust to the extent that organizational performance differs from some sort of aspirational level (Cyert and March 1963). A model of adaptive response that is based on feedback seems essential to how a team or firm motivates its choice to use its own versus others' experience in learning. Unfortunately relatively little empirical research thus far has considered this dynamic. A notable exception is Baum and Dahlin (2007) who considers how railroad accidents motivate different forms of learning. In their study, performance worse than the aspiration level motivates more exploratory search whereas performance better than the aspiration level motivates more local search. While this study and other organizational change studies have considered the ways in which unsatisfactory performance inspires experimentation to identify new ways of doing things (Greve 2003), the dynamic nature of learning a single, already-chosen practice is less understood.

III. EMPIRICAL SETTING

The empirical setting I use for the study is a Fortune 100 retail firm that made a change to its restocking practice in early 2014. The restocking practice change was a modification of the way in which employees worked to restock the shelves of each store; however the main goal of restocking stayed the same.[26] The steps of new restocking practice as well as the method for implementing the new practice were identical for all stores.

The organizational structure of the restocking team and the store management are standardized across all stores. Each store has one restocking team so a 'team' and a 'store' have the same meaning in the context of considering the restocking change and performance. For the purposes of oversight and management, stores are organized into districts and further into regions. A district is broadly defined by the locations of the stores. The stores that implemented the new restocking process during the period of study are from 34 districts in 3 regions.

The implementation of the new restocking practice followed the same pattern across districts. One store from each district was designated a pilot store, a store that would implement the process three to four weeks earlier than the rest of the stores in the district. Pilot stores were chosen as model stores for the implementation of this new restocking practice from the stores within preexisting district groupings. These pilot stores were designated prior to the announcement of the restocking change. Further, the headquarters strategy group had no specific criteria by which they chose these stores as pilots. Specifically when asked about the criteria for picking the pilot stores, a team member from the headquarters group noted that "the choice of the pilot in each region was pretty much random." In addition to the pilot store providing guidance about the new restocking practice, the headquarters provided the managers of each nonpilot store

[26] Details on the exact change to the restocking practice can be found in Lawrence (2015). For the purposes of this study, the exact revisions to the process are not necessary.

with timelines, technical manuals, and video tutorials with instructions regarding the implementation of the practice. These different tools were used to ensure that the proposed new practice and implementation were clearly communicated and uniformly understood by all stores.

Because this change to the restocking practice involved a management shift as well as a shift in the hours employees were asked to work, the headquarters wanted to ensure that the message about the restocking change was standardized. Scripts for talking to employees about the change as well as for discussing all employee transitions into and out of the department were created. The scripts were assigned to be read on certain days of announcement and in certain types of follow-up conversations. The restocking change was announced a month and a half prior to the implementation to give ample time for individual employee conversations and transitions for those employees that wanted to move to other departments or that wanted to transition into restocking from other departments.[27] Once determining the set of employees that would compose the new restocking team, all employees – new and old – were trained using the same video and informational packet content.

The practice was first implemented in pilot stores with the aim that all the managers of the restocking practice in the same district could learn from the pilot. One of the ways the nonpilot stores interacted with the pilot store was by visiting the pilot store to observe the practice firsthand. After visiting the pilot store prior to their own store implementation, the second way these stores interacted was through a weekly call with all restocking managers in the district. The purpose of this call was for the pilot store to provide instruction about the restocking

[27] Transfers into and out of the department were not incentivized in any way. The option to transfer out of the department was given to all employees since the new practice required that employees be able to work overnight hours whereas the old practice allowed for dayshifts. However, no restocking employee was required to leave the department nor was asked to leave at the time of or as a result of the change.

practice and answer questions from other stores. Once the nonpilot stores implemented the practice, the calls continued weekly with the goal of the pilot stores sharing experiences from further along in the implementation. The practice itself did not evolve over these first few months, but these calls were still viewed as a useful way for the stores to check in with the pilot and ask questions. While all stores in the district participated in each call, the only store encouraged to share successes was the pilot store. When asked why the calls were structured this way, a restocking manager said "since the pilot stores implemented the practice first, they seem like they are one step ahead in experiencing issues or tackling problems. It is really rare that anything comes up for our store that the pilot hasn't already addressed."

A final way the restocking teams in each district connected was through performance reports issued at the beginning of each week with the prior week's performance. The performance metric used is a measure of the errors that occur in the end-of-night reconciliation process to identify out-of-stock items. The issued report received by all restocking managers of each district had performance for its own store as well as all other stores, including the pilot store, in the district. A quick glance at this report gave managers a sense of how their store compared to others. Further, a restocking manager emphasized that this report gave him a sense of whether to "do more of the same or figure out what to do differently" by how well he ranked against his peers. This report was received prior to the weekly district call with the pilot store.

Several aspects of the rollout of this new restocking practice make it ideal for examining how and when stores learn from the pilot store. First, the organizational structure of the stores is predetermined as the stores were divided districts for other purposes and prior to the announcement of this restocking change. This grouping of stores creates scheduled interactions with specific stores throughout this implementation. Any other interactions with other stores in different districts happen outside of these established methods of communication and connection. Second, every effort was made to ensure that the practice itself as well as the rollout of the practice was identical across stores. While there may be small ways in which the practice differed by store that are unfortunately impossible to identify in the collected data, the likelihood of drastic and systematic differences by location is low. Lastly, the early implementation of the practice by one store in the district to act as a model for learning and also as a continual source of experience for the rest of the stores in the district provides a clear identification of outside experience for the rest of the district stores. This ability to obviously identify the reference group is rare for studies of vicarious learning and valuable for the purposes of understanding the efficacy of pilot use in practice implementation.

IV. DATA and METHODOLOGY

IV.1 DATA

The data for the study was compiled from more than 400 internal reports and files from the retail firm. Performance is measured on a weekly basis per store. In order to compile the data, I first aggregated the weekly performance measures across all stores and districts before matching this store-week data with pilot store attributes, employment data, historical information about each store, and weekly store attributes. This data includes 280 nonpilot stores and 34 pilot stores for a total of 314 stores implementing the new restocking practice. In this sample, I only include nonpilot stores in districts that also had a pilot store, a criteria which excludes 14 stores in 2 districts where no pilot store was named for the restocking practice.

The performance metric used for this study was first captured in week 2 of the rollout of the new restocking practice in the nonpilot stores. Therefore, this analysis uses data for weeks 3 through 12 of the implementation of the restocking practice in nonpilot stores which aligns with weeks 7 through 16 of the restocking practice in pilot stores. I do not include the data from the initial week of performance other than to account for it when considering lagged performance variables. I do so because this analysis focuses on learning from prior experience, and there is no way to account for prior experience in the first week of observed data. I start in week 3 and week 7 since data for week 1 for nonpilot stores as well as all data prior to week 6 for pilot stores was not captured by the headquarters or local stores. I include summary statistics and a correlation table for the main non-lagged variables in Appendix Tables 1 and 2. Overall, 2,598 observations are collected for the dependent variable; however data from more store-week pairs is used when calculating the lagged nonpilot and pilot variables of interest described below.

IV.1.1 DEPENDENT VARIABLE

I use a measure of performance called a "missed scan." This value is calculated on a weekly basis for each store as the sum of all of the missed scans that occur in a given store as the store is restocked overnight. A missed scan is generated when an error occurs in the end-of-night reconciliation portion of the new restocking practice. More specifically, at the end of each night of restocking, Monday through Friday, a member of the restocking team is responsible for using a handheld RFID scanner to scan all shelf placements in the entire store for which the SKU-specific products appear to be out-of-stock. This scanning of empty shelves occurs at the conclusion of the restocking practice, after all products have been restocked, shelves have been reorganized, and misplaced products returned to their appropriate location.

An error in the reconciliation occurs when the store computer lists an item as out-of-stock but its corresponding shelf placement is not scanned in the end-of-night process. This matching process of identifying out-of-stock items both by the computer and by scanning the shelf placement in the store is the way that items are reordered from the regional distribution center, so an error in the scanning process results in a delay in reordering products. This process of identifying out-of-stock items, both via the computer and the scanning, and creating a specific performance metric to track each store was implemented by the headquarters in response to the realization that stores often had long delays in identifying out-of-stock items. At the time of scanning shelves, the restocking team member does not know which items are out-of-stock on the store computer, so there is no opportunity to manipulate the scanning process across the entire store.

According to a few store managers, there are two common ways in which missed scans are generated. The first is a result of the way in which consumers shop for and discard items in the process of shopping. Customers often discard unwanted items throughout the store upon deciding that they do not want to purchase them. They just set down items on empty shelves rather than walking back to where they picked up the items. Further, they often set down the discarded products on open shelves, where the correct items for those shelves are out-of-stock. If these discarded items are not identified and restocked in their appropriate locations during the restocking process, there is the potential for errors to occur because shelves that should be empty are occupied by the wrong items.

The second way errors occur is when the restocking team does not complete the "packout" and "packdown" portions of the restocking practice prior to the time that the employee responsible for scanning starts moving through the aisles.[28] Because the packout and packdown have not occurred, misplaced items have not yet been tidied nor shelves straightened. Therefore, more errors may result from this miscommunication between the employees on the restocking team about the timing of restocking tasks.

This missed scan is an appropriate metric to measure performance in this setting for several reasons. First, it is the metric by which the restocking practice is evaluated by headquarters, so teams independently aim to improve their performance. Second, stores within the same district receive this metric both for themselves and all other stores in their district each week and evaluate their own performance using this reference. This benchmarking of one's own performance as compared to others is one of the ways that managers suggested they knew how they were doing. Third, this measure captures the restocking team's ability to execute the holistic restocking task – making sure that items are located in their appropriate locations in the store and that out-of-stock items are being reordered in a timely fashion.

IV.1.2 INDEPENDENT VARIABLES

IV.1.2.1 Past Missed Scan Measures

The two main variables of interest are the performance of the same store and the performance of the district pilot store in the prior week. Additional variables which capture variations on these two measures are also included.

Own Missed Scan (Last Wk): The prior performance of the store is a measure of the missed scan in the prior week. This variable captures the extent to which the prior week's performance can

[28] "Packout" and "packdown" refer to the portions of the restocking practice whereby products are moved from the storeroom to the sales floor and arranged on the appropriate shelves and extra products are either brought down from the storage shelves above or put up in the shelves above based on availability of space on the regular shelves.

be used to predict this week's performance. Since the prior week's performance is known at the start of the new week, before any restocking has occurred, the restocking manager can use it to inform decisions for the new week.

District Pilot Missed Scan (Last Wk): The pilot store in each district also measures its missed scans each week, and the prior performance of the district pilot store is the measure of that pilot store's performance in the prior week. This information is disclosed at the beginning of the week along with all other stores' performances in the district in the prior week.

Closest Pilot Missed Scan (Last Wk): Since there is one pilot store per district and that pilot store is not necessarily chosen because it is the most central store in the district, it is sometimes the case that the pilot store closest to the focal store is not actually its own district pilot. Because geographic distance matters for the transfer of knowledge (Adams and Jaffe 1996), a store may learn from the closest pilot store rather than its own, more geographically-distant pilot store. This measure captures the prior week's missed scan performance of the closest pilot store – either the stores own district pilot or another district's pilot store. Note that the nonpilot store only receives performance information about its own pilot store's performance, not nearby pilot stores in other districts.

Closest Pilot not District Pilot: I capture those stores for which the closest pilot store is not its own district pilot store using an indicator variable equal to 1 if the closest pilot is in another district and 0 otherwise.

Closest District Pilot Missed Scan (Last Wk): For those stores that are closer to another district's pilot than they are to their own pilot, I capture the performance of that closest pilot's missed scans in the prior week. This variable takes a value of zero if the closest pilot to a store is its own district pilot.

Pilot – Own Missed Scan > 0 (difference): This measure is an indicator variable capturing whether the store's own performance is lower than the district pilot's performance in the prior week. It takes a value of 1 if the focal store's performance is 'better', missing fewer errors, than its pilot and 0 otherwise.

IV.1.2.2 District Pilot Comparisons

Distance to Pilot Store: I create a measure of the distance in kilometers as-the-crow-flies between the focal store and the district pilot store. The coefficient on this measure captures if performance suffers for stores located father from the pilot for any reason.

Match Shift of Pilot: Because some stores are busier than others, there are slightly different start times and end times for different stores. Stores that have lower sales volume generally start the restocking process a few hours earlier than stores with higher sales volume. Subsequently, these stores finish much earlier in the evening than the busier stores which can be restocking until 6am. This shift in timing of the hours may cause a friction in the ability to share information with stores if the pilot store does not have the same hours or, more importantly, the same quantity of restocking to complete each night. Therefore, this measure takes a value of 1 if the focal store and pilot store have matching shifts and 0 otherwise.

Population Density of Own – Pilot Store (difference): Similarly, stores may differ in the types of customers they attract and therefore the types of purchases and orders which move products in the store. Specifically, a store manager mentioned that one of the biggest differences between stores is whether they are located in cities or suburbs. City stores have more foot traffic and smaller purchases with more deliveries than suburban stores. This difference in the types of products purchased may mean that restocking teams are facing different types of problems in reducing errors in their stores. I create a measure to capture the difference between the population density of the focal store and the pilot store in the district. I evaluate both the absolute value of the difference between population densities and the raw difference which takes positive and negative values. Greater differences between the densities may result in less relevance of the shared information as restocking managers seek to improve performance. However, this effect on sharing of information may be symmetric or asymmetric based on whether the pilot store is more urban or more rural than the focal store, hence the two ways of measuring the difference.

IV.1.2.3 District Peers Missed Scan Comparisons

Worst Store in District (Last Wk): Stores receive the performance scores of all other district stores in their own district for the prior week. Therefore, all district stores can see which store performed worst last week. I create an indicator variable to capture whether the focal store was the worst in the prior week.

Second Worst Store in District (Last Wk): Similarly, I include an indicator variable for whether a store was the second worst in the district last week. There is a significantly larger stigma associated with being the worst than being the second worst (Kuziemko et al. 2014).

Best Store in District (Last Wk): I include an indicator variable capturing the best store in the district last week. The best store in the district takes a value of 1 for the best non-pilot store in the district and zero otherwise. The best store may also be better than the district pilot in a given week (variable: *Pilot – Own Missed Scan > 0 (difference)*) in which case both of these variables take a value of 1, but these two variables are not always the same.

Avg. District Missed Scan (Last Wk): For each store-week in the data, I calculate the average performance of all other stores in the district for the prior week, excluding the focal store and the pilot store. This measure of the average performance of all the other stores in the district last week gives an indication of whether the other stores are relatively better or worse in the prior week.

Own – Lowest Missed Scan Last Week (difference): Since stores that perform much worse than other stores in their district have potentially more to improve in the following week, I capture the difference between the focal store's missed scans last week and the lowest missed scan in the district. This measure takes non-negative values only and takes a value of 0 for the store with the lowest missed scans in the district in the past week.

IV.1.3 CONTROLS

IV.1.3.1 Firm and Process Experience Measures

Even though the new restocking practice is identical in all stores, the rollout is prescribed, and the stores are completely standardized, there are differences in the experiences of the restocking teams and these differences matter. For this analysis, I include as controls the measures that were most relevant in understanding how the team experience levels impacted initial performance and subsequent team learning in Lawrence (2015).

Average Employee Experience: Using detailed employee records for all employees in the restocking department, I calculate the average years of firm experience for each restocking team each week. This measure assesses the average length of time that a member of the focal restocking team has worked for the firm in any capacity.

Supervisor Experience: Each restocking team has one supervisor who oversees the tasks of the employees and participates on the weekly call with other stores regarding the new restocking implementation. I include a measure of the years of experience that the supervisor has at the firm. Most often, the supervisor has many years of experience at the firm but does not always have the longest tenure of all the employees in the group.

Proportion of Old Guard: I control for the measure of employees that have direct or indirect experience with the old restocking practice. This measure captures the percentage of the team that was either a member of the old restocking team or a member of a different department in the focal store at the time that the restocking change announcement was made. Teams with greater proportions of these employees have been shown to start off with worse performance but improve performance more rapidly over time (Lawrence 2015).

IV.1.3.2 Store Purchase Measures

Weekly Sales Volume: Since restocking the store requires the replacement of purchased items in the store, I use a measure of the weekly sales volume as a proxy for the quantity of products that have to be replenished by the restocking team in a given week. In weeks with higher sales volume, restocking employees may have more to do during their shift which may result in a greater number of errors during the week.

Average Units per Transaction: Similarly, a measure of the average number of items purchased per customer gives a sense of the relative number of products that are moving in the store in a given week. This measure is specifically the average number of items purchased per transaction in the given week. Including this measure allows me to capture whether there are fluctuations in customer behavior week-over-week in a gives store or across stores in a single district, as fluctuations may result in differences in the restocking errors.

IV.1.3.3 Fixed Effects

As this setting is one where improvements in learning occur week-over-week, I include a week fixed effect to capture the general improvement that all stores experience in a week. The week fixed effect removes any idiosyncratic shocks that affect all stores in a given week – such as a specific sale or holiday week.

I also include district fixed effects. Districts differ dramatically in their average performance for any number of reasons. Therefore, district fixed effects rule out that the observed relationship between stores within a district is being driven by any static feature of the district such as attributes of the geographic location, district size, or district manager effects. By including both of these effects, the remaining variation in the data captures how a focal store's absolute performance relates to prior absolute and relative performance as compared to its pilot store or potentially other stores within its district.

IV.2 METHODOLOGY

IV.2.1 LEARNING FROM THE PILOT STORE

I estimate models using the *Own [store] Missed Scan* and *District Pilot Missed Scan* lagged performance variables as the main variables of interest to predict focal store-week performance:

$$MS_{idt} = \alpha + \beta_1 MS_{i,d,t-1} + \beta_2 DistrictPilotMS_{i,d,t-1} + \gamma_t + \theta_d + \sum_{k=1}^{K} \delta_k controls + \varepsilon_{idt} \quad (1)$$

where γ represents a week fixed effect, θ represents a district fixed effect, and the controls are the same as above. This model captures the *i*th store in the *d*th district at time *t*. In implementing this model, I exclude the first week of observed nonpilot missed scans from the dataset as there are no prior week's observations in the first week.

The district pilot stores are chosen by the headquarters with no specification that the stores be centrally located in their own districts. Therefore it is often the case that a store is geographically closer to another district's pilot store than to its own pilot store. This proximity to another district's pilot occurs for 26.7% of the nonpilot stores. Because knowledge has been shown to dissipate over geographic distance (Adams and Jaffe 1996), it should be the case that stores learn from the closest model store unless the groupings of stores into districts matters. I test this geographic learning effect by including the lagged performance variable for those pilot stores closest by proximity, if closer than the district pilot store in a specification as follows:

$$MS_{idt} = \alpha + \beta_1 MS_{i,d,t-1} + \beta_3 ClosestPilotMS_{i,d,t-1} + \gamma_t + \theta_d + \sum_{k=1}^{K} \delta_k controls + \varepsilon_{idt} \quad (2)$$

A nonsignificant result for the influence of these closest pilot stores would indicate that there is a nonrandom effect of the pilot in each district grouping and specifically the district pilot stores', rather than the closest pilot stores', influence store learning.

IV.2.2 MODERATORS TO LEARNING FROM THE PILOT STORE

In general, learning from a district pilot store shows up as a positive and significant relationship between the prior week's pilot store performance and this week's store performance; however in instances where the pilot store has worse performance last week than the focal store, there may be an active choice on the part of the focal store to not learn from the pilot store. If the focal store has better performance than the pilot, it may overweight its own past experience as compared to the pilot store's experience. This type of moderator of the influence of the pilot store's past performance is examined first with the aforementioned error difference between the two groups in the prior week but also using three additional differences between the pilot and focal store that may impact the weight that focal stores place on their pilot store's experiences. All of these moderators are similarly empirically specified using interactions between the variable of interest and the own store past performance or the pilot store past performance. The first relationship, whether the pilot store was better or worse than the focal store in the prior week, is modelled below:

$$MS_{idt} = \alpha + \beta_1 MS_{i,d,t-1} + \beta_2 DistrictPilotMS_{i,d,t-1} + \beta_4 \left(Pilot_{i,d,t-1} - OwnMS_{i,d,t-1} > 0\right) + \beta_5 \left(Pilot_{i,d,t-1} - OwnMS_{i,d,t-1} > 0\right) * MS_{i,d,t-1} + \beta_6 \left(Pilot_{i,d,t-1} - OwnMS_{i,d,t-1} > 0\right) * DistrictPilotMS_{i,d,t-1} + \gamma_t + \theta_d + \sum_{k=1}^{K} \delta_k controls + \varepsilon_{idt} \quad (3)$$

If the $Pilot_{i,d,t-1} - OwnMS_{i,d,t-1} > 0$ is true, the indicator variable takes a value of 1, otherwise 0. Then, the subsequent interactions of this variable with own and district pilot prior experience capture the extent to which stores put greater or less emphasis on their own or their district pilot store's performance if they are better than their pilot last week.

In addition to the raw performance difference, there are several other variables that may impact the ability or choice of a focal store to learn from its pilot store. I examine each of these in turn. These variables include the geographic distance between the focal and pilot store, the difference in the sales volume between stores as it impacts the scheduled hours of restocking work, and the difference in the general population characteristics of the area of the two stores. Each of these types of moderators is tested using the same method as above in Equation (3), with the inclusion of the moderator itself and interactions with the prior store and pilot experience. The geographic distance and population density difference are included as continuous variables whereas the difference in the scheduled hours is included as an indicator variable. I additionally test the population density difference as the absolute value of the difference. By testing both the raw difference and the absolute value of the difference, I am able to determine if there is an effect of population density difference and whether it is symmetric for stores in denser and less dense areas than the pilot stores.

IV.2.3 RELATIVE PERFORMANCE OF DISTRICT PEERS

Lastly, the extent to which stores lean on their own versus others' experiences may also be shaped by their performance as compared to the other stores in their district. Because stores receive their performance for the prior week alongside the performance for all other stores in their district on Monday morning following the prior week, store and restocking managers have the ability to compare themselves to other stores in their district before making decisions about how to change behavior the following week. This ability to compare performance with other stores may result either in differences in how stores learn from their own and pilot stores' experiences or differences in effort based on competition with peers (peer effects).

First, I test for adjustments to the own and pilot store relationship based on peer feedback. I include an indicator variable denoting the worst store in the district in the prior week. Then I interact this indicator with the missed scans of the focal store and the pilot store last week. The coefficients on these interactions indicate the extent to which the worst store in each district last week improves based on its own versus the pilot's performance. Second, I separately include an indicator variable capturing the best store in the district last week and interact it with the own store missed scans and the district pilot missed scans last week. Similar to the worst store interaction variables, the coefficients on the best store interactions capture the extent to which the focal store places weight on its own prior performance or its pilot's prior performance.

After examining differences in the relative weight placed on each type of prior performance for these best and worst performing stores, I examine two alternative ways to capture the impact of relative performance that more closely resemble peer effects. First, I examine how performance may be impacted by the average performance of other stores in the district last week. If the other stores in the district are better overall in the prior week, the measure of *Avg. District Missed Scans (Last Wk)* will be lower, and a positive and significant coefficient may indicate that better average performance last week is related to better performance of the focal store this week. Second, I examine whether performance is a function of the distance to the frontier, or how far a store is from the best performer in the district the prior week. This distance to the frontier performance may impact a store's view of its ability and need to improve performance in the next week. I examine this relationship overall as well as for the worst and second-to-worst stores in each district specifically.

These effects can be captured by including additional variables in Estimation (3). The goal of adding additional covariates to the regression is to identify whether there is any evidence that peer effects exist in these settings. The addition of these variables unfortunately does not completely rule out one effect in favor of the other or perfectly distinguish one effect from the other. However through their inclusion, I can show that the pilot store learning effect is not eliminated by these potentially simultaneously occurring peer effects.

V. RESULTS and DISCUSSION

Because stores are each assigned a pilot store within their own district that implements the new restocking practice early and that is charged with teaching the other stores how to implement the process, one major way in which districts may differ is in the pilot that stores use as a model for implementing the new practice. In Table 1, I examine the nonpilot store's current performance based on its own past performance, the past performance of its own district pilot, and the past performance of the closest pilot store by geographic proximity. In Specification (2), I estimate that the own missed scans of a store in the prior week are positively and significantly related to the missed scans experienced this week. This coefficient of 0.107 can be interpreted to mean that stores with higher missed scan values last week will also have higher missed scans this week, as compared to those stores with lower values last week. Because learning from one's own experience is incremental and, by definition, based on one's own experience, it makes sense that these values would be positively correlated. In Specification (3), I add the prior performance of the district pilot store to find that performance is also positively and significantly correlated, albeit with less than half the magnitude of the store's own experience. This finding suggests that there is some influence of the pilot store's performance on the district stores' performances – stores appear to learn from the pilot.

TABLE 1

Stores Learn From Their Own District Pilot

	(1)	(2)	(3)	(4)	(5)
	missedscan	missedscan	missedscan	missedscan	missedscan
Own Missed Scan (Last Wk)		0.107***	0.152***	0.156***	0.152***
		(0.012)	(0.012)	(0.012)	(0.012)
District Pilot Missed Scan (Last Wk)			0.064**		0.067**
			(0.031)		(0.031)
Closest Pilot Missed Scan (Last Wk)				0.027	
				(0.027)	
Indicator1: Closest Pilot not District Pilot					2.006*
					(1.036)
Indicator1 * Closest District Pilot Missed Scan (Last Wk)					-0.060
					(0.042)
Constant	22.128***	20.402***	17.586***	18.802***	17.465***
	(4.426)	(4.358)	(4.265)	(4.262)	(4.265)
Fixed Effect:	Week	Week	Week	Week	Week
	District	District	District	District	District
Controls:					
Weekly Sales (000s)	Yes	Yes	Yes	Yes	Yes
Avg Units (per trans)	Yes	Yes	Yes	Yes	Yes
Emp.Experience (yr)	Yes	Yes	Yes	Yes	Yes
Supervisor Experience (yr)	Yes	Yes	Yes	Yes	Yes
% Old Guard	Yes	Yes	Yes	Yes	Yes
% Old Guard*Emp.Exper	Yes	Yes	Yes	Yes	Yes
Obs.	2598	2598	2598	2598	2598
R-Squared	0.305	0.328	0.361	0.361	0.362

Notes: Dependent variable is missed scan errors. All columns are estimated using a fixed effect for the week and district. Dependent variable *missedscan* data is included for weeks 3 - 12 with data from week 2 missed scans only incorporated through the introduction of lagged performance variables. Lagged variables indicate performance of the respective group in the prior week and are indicated by "(Last Wk)". *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

In order to check that this learning from the pilot store is not just related to geographically local learning or influences that are not already captured by the district fixed effects, I examine the extent to which learning from the pilot store occurs by distance rather than by the district grouping. For 26.7% of the nonpilot stores, the pilot store that is closest by distance is not actually the pilot store in one's own district. Therefore, in Specification (4), I alternatively examine the influence of the closest pilot store, which in 73.3% of cases is the district pilot store and in 26.7% of the cases is another district's pilot store. I find the relationship is positive though insignificant. In Specification (5), I include a separate indicator variable and interaction variable to just capture the additional influence of the geographically proximate non-district pilot in addition to the district pilot and find no significant relationship. Therefore, I conclude that the relationship observed between a pilot and a nonpilot store in the district has something to do with the district itself rather than a merely geographic proximity to any pilot store.

Once determining that stores seem to place weight on both their own and their pilot's prior performance with the new restocking process, I examine how performance feedback moderates the weight placed on these learning relationships. The first type of performance feedback I examine is whether the nonpilot store had better performance than the pilot store in the prior week. These results are shown in Table 2. Specification (1) repeats the results from Table 1 Specification (3). In Specification (2), I include an indicator variable equal to 1 if the nonpilot store's missed scans in the prior week were lower than the pilot store's missed scans. The nonpilot store performance is better than (lower than) the pilot store 38.7% of the time.[29] I find that performance in the following week is generally significantly improved if the nonpilot store was better than the pilot store in the prior week. However, Specification (3), which includes the interaction of the indicator variables with the prior week's performance for both the nonpilot and pilot store, is more interesting. I find that stores that perform better than the pilot in the prior week place a greater emphasis on their own performance, as evidenced by the positive and significant coefficient on the *Indicator2 * Own Missed Scan (Last Wk)* interaction. In contrast, the nonpilot store places less emphasis on the experiences of the pilot store, as evidenced by the negative and significant coefficient on the *Indicator2 * District Pilot Missed Scan (Last Wk)* interaction. In fact, when estimating a linear combination of the coefficient on last week's pilot

[29] The nonpilot store performance indicator takes a value of 1 only if both the pilot store and own store missed scans last week are not missing and the pilot store missed scans are higher than the nonpilot store scans. These conditions are met in 1,006 out of a total of 2,598 instances.

missed scan and the interaction term, I find that the combination of these values is not distinguishable from zero (0.330 + -0.349 = -0.019), meaning that nonpilot stores place no significant weight on the pilot store's prior performance in weeks following those where they perform better – rather only weighting their own past experiences.[30] This finding suggests that restocking managers do indeed look at performance metrics before deciding whether to use their own or outside experiences to improve in the coming week.

TABLE 2

Learning Moderated by Pilot Performance

	(1) missedscan	(2) missedscan	(3) missedscan
Own Missed Scan (Last Wk)	0.152***	0.107***	0.078***
	(0.012)	(0.013)	(0.014)
District Pilot Missed Scan (Last Wk)	0.064**	0.165***	0.330***
	(0.031)	(0.033)	(0.048)
Indicator2: Pilot - Own Missed Scan > 0		-6.885***	-4.666***
		(0.800)	(0.978)
Indicator2 * Own Missed Scan (Last Wk)			0.322***
			(0.067)
Indicator2 * District Pilot Missed Scan (Last Wk)			-0.349***
			(0.060)
Constant	17.586***	20.016***	17.111***
	(4.265)	(4.214)	(4.214)
Fixed Effect:	Week District	Week District	Week District
Controls:			
Weekly Sales (000s)	Yes	Yes	Yes
Avg Units (per trans)	Yes	Yes	Yes
Emp.Experience (yr)	Yes	Yes	Yes
Supervisor Experience (yr)	Yes	Yes	Yes
% Old Guard	Yes	Yes	Yes
% Old Guard*Emp.Exper	Yes	Yes	Yes
Obs.	2598	2598	2598
R-Squared	0.361	0.379	0.387

Notes: Dependent variable is missed scan errors. All columns are estimated using a fixed effect for the week and district. Dependent variable *missedscan* data is included for weeks 3 - 12 with data from week 2 missed scans only incorporated through the introduction of lagged performance variables. Lagged variables indicate performance of the respective group in the prior week and are indicated by "(Last Wk)". *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

[30] The joint coefficient was estimated using the *lincom* command in Stata which provides estimates of the linear combination of coefficients along with a single standard error for the combined estimate.

In addition to the raw performance comparison with the pilot store, a nonpilot store may experience contextual barriers to learning from the pilot store. These other main moderators between the experiential and vicarious learning choices of stores are considered in Table 3. First, in Specification (2) of Table 3, I consider the overall impact of geographic distance from the pilot store on the performance of the store each week. I find no evidence that stores farther from the pilot store in the district perform any worse than those closer to the pilot store as the coefficient on distance is insignificant. Also, stores do not seem to place any greater or less weight on their own experience versus the pilot's experience due to distance (Specification (3)). This null result also suggests that there is no evidence that more isolated or remote stores in the district are somehow shirking on their effort because of perhaps less frequent or rigorous monitoring or interaction with other stores.

TABLE 3

Contextual Differences Impact Type of Learning

	(1) missedscan	(2) missedscan	(3) missedscan	(4) missedscan	(5) missedscan	(6) missedscan
Own Missed Scan (Last Wk)	0.078***	0.078***	0.057***	0.196***	0.100***	0.102***
	(0.014)	(0.014)	(0.016)	(0.032)	(0.019)	(0.015)
District Pilot Missed Scan (Last Wk)	0.330***	0.330***	0.317***	0.284***	0.349***	0.287***
	(0.048)	(0.048)	(0.051)	(0.061)	(0.052)	(0.048)
Indicator2: Pilot - Own Missed Scan > 0	-4.666***	-4.666***	-4.504***	-4.274***	-4.694***	-4.501***
	(0.978)	(0.978)	(0.978)	(0.982)	(0.976)	(0.974)
Indicator2 * Own Missed Scan (Last Wk)	0.322***	0.322***	0.314***	0.305***	0.312***	0.274***
	(0.067)	(0.067)	(0.067)	(0.067)	(0.067)	(0.067)
Indicator2 * District Pilot Missed Scan (Last Wk)	-0.349***	-0.349***	-0.334***	-0.331***	-0.331***	-0.293***
	(0.060)	(0.060)	(0.060)	(0.060)	(0.060)	(0.061)
Distance to Pilot Store (km)		0.000	-0.013			
		(0.009)	(0.012)			
Own Missed Scan (Last Wk) * Distance to Pilot			0.001***			
			(0.000)			
District Pilot Missed Scan (Last Wk) * Distance to Pilot			-0.000			
			(0.001)			
Indicator3: Match the Shift of Pilot (Day or Night)				2.047**		
				(0.965)		
Indicator3 * Own Missed Scan (Last Wk)				-0.130***		
				(0.032)		
Indicator3* District Pilot Missed Scan (Last Wk)				0.038		
				(0.048)		
Abs(Difference in Population Density of Own - Pilot Store)					0.142	
					(0.223)	
Abs(Difference in Population Density) * Own Missed Scan (Last Wk)					-0.008*	
					(0.004)	
Abs(Difference in Population Density) * District Pilot Missed Scan (Last Wk)					-0.016*	
					(0.008)	
Difference in Population Density of Own - Pilot Store						-0.490**
						(0.202)
Difference in Population Density * Own Missed Scan (Last Wk)						0.016***
						(0.003)
Difference in Population Density * District Pilot Missed Scan (Last Wk)						-0.003
						(0.007)
Constant	17.111***	17.112***	18.234***	15.642***	16.568***	16.660***
	(4.214)	(4.222)	(4.232)	(4.244)	(4.239)	(4.202)
Fixed Effect:	Week District	Week District	Week District	Week District	Week District	Week District
Controls:						
Weekly Sales (000s)	Yes	Yes	Yes	Yes	Yes	Yes
Avg Units (per trans)	Yes	Yes	Yes	Yes	Yes	Yes
Emp.Experience (yr)	Yes	Yes	Yes	Yes	Yes	Yes
Supervisor Experience (yr)	Yes	Yes	Yes	Yes	Yes	Yes
% Old Guard	Yes	Yes	Yes	Yes	Yes	Yes
% Old Guard*Emp.Exper	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	2598	2598	2598	2598	2598	2598
R-Squared	0.387	0.387	0.390	0.392	0.390	0.394

Notes: Dependent variable is missed scan errors. All columns are estimated using a fixed effect for the week and district. Dependent variable *missedscan* data is included for weeks 3 - 12 with data from week 2 missed scans only incorporated through the introduction of lagged performance variables. Lagged variables indicate performance of the respective group in the prior week and are indicated by "(Last Wk)". *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

While geographic distance does not impact the pilot-nonpilot prior performance relationship, the match between restocking store hours worked between the two stores does have an impact on the types of learning each week. Specification (4) includes an indicator variable (*Indicator3: Match the Shift of Pilot (Day or Night)*) for whether the shift hours match exactly between the pilot and nonpilot stores. Shift hours can vary slightly as to whether they start in early or late evening (called "day" and "night" shifts). The coefficient on the indicator variable captures the effect of shift hour matches. I interact this indicator with the own missed scans and pilot missed scans in the prior week to find that stores assign proportionately less weight to their own store performance (-0.130) if the shifts match and proportionately more, but not significantly so, to the pilot store performance (0.038).

Lastly, I consider the impact of differences in population density between the pilot store location and the nonpilot store location in Specifications (5) and (6). Population density differences, or rather differences in "city" versus "suburban" stores, were cited as a major difference between store customer behavior and store operations by store managers. Differences along this dimension are associated with differences in the types of products purchased by store, the order method, and the delivery method. Therefore, it is reasonable to expect that mismatches in population density between pilot store and nonpilot store result in less applicability of pilot store experiences to nonpilot stores. In Specification (5), I consider the impact of an absolute difference between the population density of the focal store and that of the pilot. I find no raw effect but slightly reduced positive correlation between the prior scans of both the focal store and the pilot store's performance last week and the focal store's performance this week.

It is possible that the effect of the population density difference is not symmetric. Stores that are larger than the pilot may feel like there is less to learn from a more rural store as rural stores tend to have fewer of the complicating factors like customer deliveries and online orders. Therefore, in Specification (6), I include a measure of the raw difference in population densities between the focal store and the pilot store. Here, I see that stores that are in more urban areas than the pilot tend to be increasingly correlated with their own performance. There is a negative, though small and insignificant, decrease in the correlation with the pilot store's performance last week for these larger stores. Overall the results in Table 3 suggest there is some significant effect of similarity between the focal store and the pilot store on the relationship between prior performance and current performance. If focal stores more closely resemble the pilot store, they place less weight on their own prior performance.

In addition to the pilot-nonpilot store relationship, performance feedback as contextualized by the other district stores may also impact the way in which a nonpilot store improves new practice performance over time. Specifically, because stores see the performance of all other stores in their districts at the start of the new week, stores may either glean information about their own versus others' performances that impacts the way they improve or they may be differently motivated to improve due to their specific peer group. While it is hard to explicitly tease apart the effect of peer performance on motivation from its effect on learning, I examine the potential peer relationship in several ways to confirm that the learning relationship remains even amidst perhaps simultaneously occurring motivation-based peer effects.

First, I examine the impact of occupying either the prominent last place or prominent first place position in the district performance report in the prior week in Table 4. In particular, I examine how being first or last affects the amount of weight that stores seem to allocate to their own versus their pilot's prior performance. In Specification (2), I consider the raw effect of being worst by including an indicator variable for the worst store in each district each week as well as the interaction with the own store and pilot store prior performance. The positive and significant coefficient on the raw worst store variable (*Indicator4*) suggests that stores often persist to some extent week-over-week in their poor performance, performing worse than average in the following week as well. Stores with the worst performance in the district have no correlation with their own performance last week as the sum of the coefficients on own missed scan last week (0.361) and worst store own missed scan last week (-0.374) is statistically indistinguishable from zero (Specification (2)). However, these worst stores seem to more heavily weight the district pilot's performance last week, with a positive and significant coefficient on the *Indicator4 * District Pilot Missed Scan (Last Wk).*

In Specification (3), I consider the impact of being the best nonpilot store in the district last week, both capturing the raw and interaction effects with prior own and pilot performances. With *Indicator5*, I find that the best store in the district last week tends to perform significantly better this week as well. Interestingly, the best store in the district last week seems to have a negative and significant relationship with its own performance last week (0.062 + -0.335) as compared to other stores which maintain their positive and significant relationship with own performance. There is no significant difference in the weight that these best stores put on their pilot's prior performance as compared to other stores in their district unless they perform better than their pilot in the prior period – which has already been noted to change this relationship (see Table 2) and is confirmed in Specification (3). Together, the way that these positional rankings relate to the weight placed on own versus pilot store prior performance broadly support a notion that stores seem to learn from themselves and their pilot and that positional rankings influence the relative types of learning rather than eliminating the learning effect.

TABLE 4

Learning Effects Differ for Relatively Better and Worse Performers

	(1) missedscan	(2) missedscan	(3) missedscan
Own Missed Scan (Last Wk)	0.078***	0.361***	0.062***
	(0.014)	(0.032)	(0.014)
District Pilot Missed Scan (Last Wk)	0.330***	-0.003	0.335***
	(0.048)	(0.059)	(0.047)
Indicator2: Pilot - Own Missed Scan > 0	-4.666***	-0.948	-2.826***
	(0.978)	(0.998)	(1.061)
Indicator2 * Own Missed Scan (Last Wk)	0.322***	0.024	0.264***
	(0.067)	(0.071)	(0.076)
Indicator2 * District Pilot Missed Scan (Last Wk)	-0.349***	-0.024	-0.345***
	(0.060)	(0.069)	(0.066)
Indicator4: Worst Store in District (Last Wk)		13.794***	
		(1.432)	
Indicator4 * Own Missed Scan (Last Wk)		-0.374***	
		(0.034)	
Indicator4 * District Pilot Missed Scan (Last Wk)		0.295***	
		(0.073)	
Indicator5: Best Store in District (Last Wk)			-5.120***
			(1.166)
Indicator5 * Own Missed Scan (Last Wk)			-0.335***
			(0.095)
Indicator5 * District Pilot Missed Scan (Last Wk)			0.013
			(0.066)
Constant	17.111***	10.101**	16.464***
	(4.214)	(4.125)	(4.159)
Fixed Effect:	Week District	Week District	Week District
Controls:			
Weekly Sales (000s)	Yes	Yes	Yes
Avg Units (per trans)	Yes	Yes	Yes
Emp.Experience (yr)	Yes	Yes	Yes
Supervisor Experience (yr)	Yes	Yes	Yes
% Old Guard	Yes	Yes	Yes
% Old Guard*Emp.Exper	Yes	Yes	Yes
Obs.	2598	2598	2598
R-Squared	0.387	0.425	0.404

Notes: Dependent variable is missed scan errors. All columns are estimated using a fixed effect for the week and district. Dependent variable *missedscan* data is included for weeks 3 - 12 with data from week 2 missed scans only incorporated through the introduction of lagged performance variables. Lagged variables indicate performance of the respective group in the prior week and are indicated by "(Last Wk)". *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

In Table 5, I begin to consider the effect of peer performance, separate from the pilot store performance, as a predictor of the focal store's performance. First, in Specification (1), I include a measure of the average missed scans of all other stores in the district last week except the focal store and the pilot store. A significant coefficient on this variable would suggest that the aggregate performance of other stores in the district last week has an impact on performance of the focal store this week. However, the insignificant coefficient is taken as one sign that the aggregate effect of one's peers last week is not significantly related to performance this week.

TABLE 5

Performance as a Function Relative Performance to Peers

	(1) missedscan	(2) missedscan	(3) missedscan	(4) missedscan
Own Missed Scan (Last Wk)	0.181***	0.039	0.037	0.033
	(0.012)	(0.057)	(0.054)	(0.054)
District Pilot Missed Scan (Last Wk)	0.101***	0.076**	0.024	0.026
	(0.031)	(0.031)	(0.030)	(0.030)
Avg District Missed Scan (Last Wk)	-0.012			
	(0.021)			
Own Missed Scan - Lowest District Missed Scan (Last Wk)		0.122**	0.401***	0.393***
		(0.060)	(0.061)	(0.065)
Indicator4: Worst Store in District (Last Wk)			17.276***	17.523***
			(1.239)	(1.249)
Indicator4 * (Own Missed Scan - Lowest District Missed Scan (Last Wk))			-0.442***	-0.431***
			(0.028)	(0.036)
Indicator6: Second Worst Store in District (Last Wk)				3.837***
				(1.468)
Indicator6 * (Own Missed Scan - Lowest District Missed Scan (Last Wk))				-0.063
				(0.052)
Constant	17.128***	17.629***	9.332**	9.124**
	(4.176)	(4.262)	(4.067)	(4.067)
Fixed Effect:	Week District	Week District	Week District	Week District
Controls:				
Weekly Sales (000s)	Yes	Yes	Yes	Yes
Avg Units (per trans)	Yes	Yes	Yes	Yes
Emp.Experience (yr)	Yes	Yes	Yes	Yes
Supervisor Experience (yr)	Yes	Yes	Yes	Yes
% Old Guard	Yes	Yes	Yes	Yes
% Old Guard*Emp.Exper	Yes	Yes	Yes	Yes
Obs.	2598	2598	2598	2598
R-Squared	0.388	0.362	0.428	0.429

N*otes: D*ependent variable is missed scan errors. All columns are estimated using a fixed effect for the week and district. Dependent variable *missedscan* data is included for weeks 3 - 12 with data from week 2 missed scans only incorporated through the introduction of lagged performance variables. Lagged variables indicate performance of the respective group in the prior week and are indicated by "(Last Wk)". *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

Next, in Specification (2), I include a measure of how many more missed scans a store has than the best store in its district and find there is a positive and significant relationship between this value in the prior week and performance this week. In other words, poor performance seems to persist week-over-week. Better performing stores seem to have better performance the following week and worse performing stores have worse performance the following week. However, the worst store in the district last week seems to behave differently (see Specification (3)). In contrast to the general persistence of worse performance week-over-week, the worst store in the district seems to have no relationship between how far its prior performance is from the frontier last week and its performance this week. The sum of the two coefficients (0.401 + -0.442) is not significantly different from zero for these worst stores. Further, this effect is only present for the worst store in the district, not the second worst store, as seen in Specification (4) with the inclusion of *Indicator6* and the interaction of this indicator with the distance to the frontier measure. The second worst store in any district last week seems to have the same persistent poor performance relative to its position last week. From these findings, it seems as if there is some evidence to suggest that the worst stores in each district are perhaps motivated by their performance relative to peers to perform better in the following weeks; however there does not seem to be an aggregate relationship between peer's relative performance and subsequent improvement in focal store performance.

VI. DISTINGUISHING LEARNING and PEER EFFECTS MECHANISMS

Thus far, the empirical tests and results have assumed that the mechanism driving the relationship between pilot store performance and district performance is stores learning from their pilot stores with only some analysis of the potential effects of relative peer performance. It may be the case that the relationship between the performance of pilot stores and nonpilot stores is due to mainly to an alternative mechanism. The most likely alternative has already been mentioned and is that the division of stores into districts creates a set of peer stores that compete with one another to achieve better performance. Distinguishing between peer effects and learning is challenging which is why these two effects are often considered separately and with some necessary simplifying assumptions to distinguish each case (e.g. Chan, Li, and Pierce 2014a; Chan, Li, and Pierce 2014b).

Thus far, I have provided evidence of the impact of prior own store and pilot store performance on current performance and shown that this relationship persists even when considering aggregate district effects such as the lagged average performance of other stores in the district. Also, I have shown that there seems to be some persistence in performance week-over-week related to a store's distance to the frontier performance of its district. With the exception of the worst store in the district, a store's relative performance compared to its best district peer in a prior week is associated with increased missed scans in the following week. However, these results do not conclusively differentiate whether the pilot store actually matters because it is the pilot store and stores learn from it or because it is another example of a peer and perhaps the most prominent peer. Therefore, I consider an additional set of specifications as an alternative test of the pilot-nonpilot store relationship and the peer effects relationship. This test relies on a notion that learning occurs in proportion to the amount to be learned, or the cardinal performance improvement capacity, while peer effects most directly are a result of the ordinal rank of performance. In other words, I make an assumption that learning is more likely to occur when there is a greater disparity between the pilot store and the focal nonpilot store performance last week. In contrast, peer effects or differences in motivation to improve are more likely to occur based on the focal nonpilot store's performance rank in its district – best, worst, second-to-worst, etc. Kuziemko et al. (2014) find that individuals are "last place averse" which would suggest that, at the very least, a peer effect should result in greater improvement for those stores ranking worst in their district in the prior week.

In Table 6, I examine the week-over-week improvement in missed scans for the nonpilot store as a function of its cardinal performance improvement capacity and its ordinal rank of performance in its district. In Specification (1), I include a measure of how much better the pilot store was than the focal nonpilot store in the prior week, if it was better. This measure is a simple difference but takes a value of zero if the nonpilot store outperformed or performed equally as well as the pilot store last week. To capture stores that outperformed their pilot last week, I also include the same indicator variable from previous regressions, *Indicator2*. I find a positive and significant relationship between week-over-week improvement and the improvement capacity. This finding suggests that stores with potentially more to learn from the pilot do reduce missed scans by a greater amount in the following week. Next, in Specifications (2) – (4), I include indicator variables to examine how the rank of a store in its district contributes to its week-over-week improvement. When testing the worst store in the district (Specification (2)), the second worst store in the district (Specification (3)), and the best store in the district (Specification (4)), I find no significant relationship between the district rank of performance in the prior week and subsequent improvement. In all of these cases, I am controlling for district and week fixed effects; however the results are robust when removing the district fixed effects. Together, the significant relationship based on the cardinal difference from the pilot store and insignificant relationship of the ordinal rank of the stores in the district suggest that learning from the pilot store is most likely at contributing factor to how stores improve performance whereas peer effects may or may not be significant. These results are not able to conclusively rule out evidence of peer effects, and, of course, the prior evidence in Table 5 showing that worst place stores improve is already a form of peer effect. In the following section, I outline some potential next steps to more conclusively disentangle these two mechanisms.

TABLE 6

Week-over-Week Improvement in Missed Scans

	(1) WoW Improvement	(2) WoW Improvement	(3) WoW Improvement	(4) WoW Improvement
Own - Pilot Missed Scan (Last Wk)	0.029***	0.029***	0.028***	0.028***
	(0.006)	(0.007)	(0.007)	(0.006)
Indicator2: Pilot - Own Missed Scan > 0	1.268***	1.271***	1.342***	1.370***
	(0.319)	(0.323)	(0.332)	(0.341)
Indicator4: Worst Store in District (Last Wk)		0.022	0.143	
		(0.493)	(0.510)	
Indicator6: Second Worst Store in District (Last Wk)			0.408	
			(0.436)	
Indicator5: Best Store in Distirct (Last Wk)				-0.337
				(0.399)
Constant	-39.390***	-39.397***	-39.440***	-39.525***
	(1.969)	(1.975)	(1.975)	(1.975)
Fixed Effect:	Week	Week	Week	Week
	District	District	District	District
Controls:				
Weekly Sales (000s)	Yes	Yes	Yes	Yes
Avg Units (per trans)	Yes	Yes	Yes	Yes
Emp.Experience (yr)	Yes	Yes	Yes	Yes
Supervisor Experience (yr)	Yes	Yes	Yes	Yes
% Old Guard	Yes	Yes	Yes	Yes
% Old Guard*Emp.Exper	Yes	Yes	Yes	Yes
Obs.	2598	2598	2598	2598
R-Squared	0.706	0.706	0.706	0.706

Notes: Dependent variable is Week-over-Week Improvement in missed scans by store. In other words, it is the first difference of the missed scans. Dependent variable *WoW Improvement* data is included for weeks 3 - 12 with data from week 2 missed scans only incorporated via the calculation of the first week's improvement. *Own - Pilot Missed Scan (Last Wk)* captures how much worse the focal store was than its district pilot store last week, taking positive values when the focal store is worse than the pilot store a value of zero if it is better than the pilot store last week. Fixed effects are included at the week and district level. *** denotes 1% significance, ** denotes 5% significance, and * denotes 10% significance.

VII. CONCLUSION

These results show evidence of a significant effect of pilot use on the way that stores perform a new practice. Stores seem to learn and be influenced locally, rather than geographically, by their pilot stores which suggests that interunit transfer of information does occur vertically from the pilot down to unit rather than directly and horizontally between proximate units in hierarchically organized multi-unit firms (Schulz 2003). This finding is further supported by a lack of evidence that geographic proximity to the pilot store in the district impacts performance of nonpilot stores. The creation of the district, while admittedly composed of already somewhat proximate stores, seems to create a focused group for pilot store learning.

Further, while previous studies have focused on one mode of learning, it is evident from this study that learning modes – learning from own and from others' experiences – are interdependent and simultaneous. Teams learn both from their own experience and from the pilots' experiences with the new practice over time. Performance feedback about the relative performance of teams compared to their district impacts the extent to which teams look outside their own experiences to learn over time. Contextual similarity, as measured in this paper by restocking team shift overlap and store population density similarity, also seem to affect the extent to which restocking teams rely on their own versus their pilot store's prior performance when learning.

While most of this paper has focused on learning from the pilot store as the proposed function of the pilot store and mechanism by which the pilot and nonpilot stores performances are related, the creation of a district reference group and sharing of performance metrics weekly for all stores in that group does suggest another viable alternative mechanism. This alternative mechanism may be differences in motivation or effort to learn based on the peer competitive effects created through district groupings. I have provided initial evidence in Table 6 that suggests week-over-week improvement in missed scans is not related to the ordinal positioning of stores within a district though this type of relationship might be expected of stores competing with their peers based on their prior performance. However, more work is needed to conclusively distinguish the peer effects model from that of the proposed learning model. As an example of a next step, I plan to explore whether pilot store prior performance itself is distinctive or whether the performance of a randomly chosen store in the district might appear as strongly and significantly related to current performance of other district stores. If this relationship between prior performance of the pilot and current performance of the nonpilot only exists for the chosen pilot, it provides additional evidence that the specific designation of a pilot store does matter.

An ideal experiment testing the efficacy and impact of pilot stores would have randomly allocated some districts to have pilots and others to not have pilots. Similarly, an ideal test of peer effects would have allowed for some districts to see their peers' performances and participate on joint weekly calls together while other districts learned in isolation. Unfortunately, this setting provides neither of those settings.

Nevertheless, this setting does provide a rare opportunity to examine how practice transfer and implementation occur using real firm data at a relatively granular level. Even if these two proposed mechanisms, learning and peer effects, cannot be completely distinguished, this examination of practice transfer using pilots can still shed light on how firms transition to new practices which is a crucial component of understanding how firms are able to maintain a competitive advantage over time in changing environments.

BIBLIOGRAPHY

Adams, James D, and Adam B Jaffe. 1996. "Bounding the Effects of R & D : An Investigation Using Matched Establishment-Firm Data." *The RAND Journal of Economics* 27 (4): 700–721.

Argote, Linda. 2013. *Organizational Learning: Creating, Retaining and Transferring Knowledge*. 2nd Edition. New York: Springer.

Argote, Linda, S. L. Beckman, and Dennis Epple. 1990. "The Persistence and Transfer of Learning in Industrial Settings." *Management Science* 36 (2): 140–154.

Argote, Linda, and Paul Ingram. 2000. "Knowledge Transfer: A Basis for Competitive Advantage in Firms." *Organizational Behavior and Human Decision Processes* 82 (1): 150–169.

Baum, J. a. C., and K. B. Dahlin. 2007. "Aspiration Performance and Railroads' Patterns of Learning from Train Wrecks and Crashes." *Organization Science* 18 (3): 368–385.

Benkard, C. Lanier. 2000. "Learning and Forgetting: The Dynamics of Aircraft Production." *American Economic Review* 90 (4): 1034–1054.

Bingham, Christopher B., and Jason P. Davis. 2012. "Learning Sequences: Their Existence, Effect and Evolution." *Academy of Management Journal* 55 (3): 611–641.

Bloom, Nicholas, B Eifert, A Mahajan, D McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India." *The Quarterly Journal of Economics* 128 (1): 1–51.

Bresman, Henrik. 2013. "Changing Routines: A Process Model of Vicarious Group Learning in Pharmaceutical R&D." *Academy of Management Journal* 56 (1): 35–61.

Chan, Tat Y, Jia Li, and Lamar Pierce. 2014a. "Learning from Peers : Knowledge Transfer and Sales Force Productivity Growth." *Marketing Science* 33 (4): 463–484.

———. 2014b. "Compensation and Peer Effects in Competing Sales Teams." *Management Science* 60 (8): 1965–1984.

Chandler, Alfred. 1977. *The Visible Hand: The Managerial Revolution in American Business*. Cambridge, MA: Harvard University Press.

Chang, Myong-Hun, and Joseph E. Harrington. 2000. "Centralization vs. Decentralization in a Multi-Unit Organization: A Computational Model of a Retail Chain as a Multi-Agent Adaptive System." *Management Science* 46 (11): 1427–1440.

Cyert, Richard M., and James G. March. 1963. *A Behavioral Theory of the Firm*. Englewood Cliffs, N.J.: Prentice-Hall.

Darr, Eric D., Linda Argote, and Dennis Epple. 1995. “The Acquisition, Transfer, and Depreciation of Knowledge in Service Organizations: Productivity in Franchises.” *Management Science* 41 (11): 1750–1762.

Gibbons, Robert, and Rebecca Henderson. 2012. “What Do Managers Do? Exploring Persistent Performance Differences among Seemingly Similar Enterprises.” *Harvard Business School Working Paper 13-020*.

Greve, Henrich R. 1998. “Performance Aspirations, and Risky Change Organizational Change.” *Administrative Science Quarterly* 43 (1): 58–86.

———. 2003. *Organizational Learning from Performance Feedback: A Behavioral Perspective on Innovation and Change*. Cambridge, UK: Cambridge University Press.

Gupta, A, DG Hoopes, and AM Knott. 2014. “Redesigning Routines for Replication.” *Strategic Management Journal* 36 (6): 851–871.

Haunschild, Pamela R, and Anne S Miner. 1997. “Modes of Interorganizational Imitation: The Effects of Outcome Salience and Uncertainty.” *Administrative Science Quarterly* 42 (3): 472–500.

Helfat, Constance E., and Margaret A. Peteraf. 2003. “The Dynamic Resource-Based View: Capability Lifecycles.” *Strategic Management Journal* 24 (10): 997–1010.

Helfat, Constance E., and Sidney G. Winter. 2011. “Untangling Dynamic and Operational Capabilities: Strategy for the (N) Ever-Changing World.” *Strategic Management Journal* 32 (11): 1243–1250.

Huber, GP. 1991. “Organizational Learning: The Contributing Processes and the Literatures.” *Organization Science* 2 (1): 88–115.

Ingram, Paul, and Joel A C Baum. 1997. “Opportunity and Constraint: Organizations’ Learning from the Operating and Competitive Experience of Industries.” *Strategic Management Journal* 18 (Special Issue: Organizational and Competitive Interactions): 75–98.

Jensen, Robert, and Gabriel Szulanski. 2004. “Stickiness and the Adaptation of Organizational Practices in Cross-Border Knowledge Transfers.” *Journal of International Business Studies* 35 (6): 508–523.

———. 2007. “Template Use and the Effectiveness of Knowledge Transfer.” *Management Science* 53 (11): 1716–1730.

Kalnins, Arturs, and Kyle J. Mayer. 2004. “Franchising, Ownership, and Experience: A Study of Pizza Restaurant Survival.” *Management Science* 50 (12): 1716–1728.

Knott, Anne Marie. 2003. “The Organizational Routines Factor Market Paradox.” *Strategic Management Journal* 24 (10): 929–943.

Kogut, Bruce, and Udo Zander. 1992. "Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology." *Organization Science* 3 (3): 383–397.

Kuziemko, Ilyana, Ryan W Buell, Taly Reich, and Michael I Norton. 2014. ""Last-Place Aversion'': Evidence and Redistributive Implications." *The Quarterly Journal of Economics* 129 (1): 105–149.

Lawrence, Megan. 2015. "Taking Stock of the Ability to Change: Prior Experience, Competency Traps, and Learning-by-Doing." *Working Paper*.

Levinthal, Daniel A, and James G. March. 1993. "The Myopia of Learning." *Strategic Management Journal* 14: 95–112.

Levitt, Barbara, and James G. March. 1988. "Organizational Learning." *Annual Review of Sociology* 14 (1): 319–338.

March, James. 1991. "Exploration and Exploitation in Organizational Learning." *Organization Science* 2 (1): 71–87.

———. 2010. *The Ambiguities of Experience*. Ithaca, N.Y.: Cornell University Press.

Nelson, Richard, and Sidney G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, MA: The Belknap Press of Harvard University Press.

Parmigiani, Anne, and SS Holloway. 2011. "Actions Speak Louder than Modes: Antecedents and Implications of Parent Implementation Capabilities on Business Unit Performance." *Strategic Management Journal* 32 (5): 457–485.

Pisano, Gary P., Richard M.J. Bohmer, and Amy C. Edmondson. 2001. "Organizational Differences in Rates of Learning: Evidence from the Adoption of Minimally Invasive Cardiac Surgery." *Management Science* 47 (6): 752–768.

Prahalad, C K, and Gary Hamel. 1990. "The Core Competence of the Corporation." Edited by Dietger Hahn and Bernard Taylor. *Harvard Business Review* 68 (3): 79–91.

Reagans, Ray, Linda Argote, and Daria Brooks. 2005. "Individual Experience and Experience Working Together: Predicting Learning Rates from Knowing Who Knows What and Knowing How to Work Together." *Management Science* 51 (6): 869–881.

Rivkin, Jan W. 2000. "Imitation of Complex Strategies." *Management Science* 46 (6): 824–844.

———. 2001. "Reproducing Knowledge: Replication without Imitation at Moderate Complexity." *Organization Science* 12 (3): 274–293.

Schulz, Martin. 2003. "Pathways of Relevance: Exploring Inflows of Knowledge into Subunits of Multinational Corporations." *Organization Science* 14 (4): 440–459.

Schwab, Andreas. 2007. "Incremental Organizational Learning from Multilevel Information

Sources: Evidence for Cross-Level Interactions.” *Organization Science* 18 (2): 233–251.

Sorenson, Olav, and Jesper B. Sorensen. 2001. “Finding the Right Mix: Franchising, Organizational Learning, and Chain Performance.” *Strategic Management Journal* 22 (6-7): 713–724.

Szulanski, Gabriel. 1996. “Exploring Internal Stickiness: Impediments to the Transfer of Best Practice Within the Firm.” *Strategic Management Journal* 17 (Winter Special Issue): 27–43.

Szulanski, Gabriel, and Sidney G. Winter. 2002. “Getting It Right the Second Time.” *Harvard Business Review* 80: 62–71.

Teece, David J. 2007. “Explicating Dynamic Capabilitites: The Nature and Microfoundations of (Sustainable) Enterprise Performance.” *Strategic Management Journal* 28 (13): 1319–1350.

Williams, Charles. 2007. “Transfer in Context: Replication and Adaptation in Knowledge Transfer Relationships.” *Strategic Management Journal* 28 (9): 867–889.

Winter, Sidney G., and Gabriel Szulanski. 2001. “Replication as Strategy.” *Organization Science* 12 (6): 730–743.

Wright, Theodore Paul. 1936. “Learning Curve.” *Journal of the Aeronautical Sciences* 3 (1): 122–128.

Zollo, Maurizio, and Sidney G. Winter. 2002. “Deliberate Learning and the Evolution of Dynamic Capabilities.” *Organization Science* 13 (3): 339–351.

APPENDIX

APPENDIX TABLE 1

Summary Statistics

Variable	Observations	Mean	Std. Dev.	Min	Max
missedscan	2598	15.7071	19.1984	0	210
Weekly Sales (000s)	2598	733.9861	274.8351	217.509	1874.356
Avg Units (per trans)	2598	6.3738	0.6738	4.320	10.677
Emp.Experience (yr)	2598	3.2592	1.5076	0.714	10.497
Supervisor Experience (yr)	2598	7.8103	4.8501	0	26.808
% Old Guard	2598	0.5265	0.1563	0.143	1
Indicator1: Closest Pilot not District Pilot	2598	0.2698	0.4440	0	1
Indicator2: Pilot - Own Missed Scan > 0	2598	0.3803	0.4856	0	1
Distance to Pilot Store (km)	2598	42.8144	43.9675	2.142	235.067
Indicator3: Match the Shift of Pilot (Day or Night)	2598	0.6594	0.4740	0	1
Abs(Difference in Population Density of Own - Pilot Store)	2598	1.9872	2.5270	0.003	13.532
Difference in Population Density of Own - Pilot Store	2598	-0.0441	3.2147	-12.685	13.532
Own Missed Scan - Lowest District Missed Scan (Last Wk)	2598	14.5131	26.0235	0	609
Indicator4 * (Own Missed Scan - Lowest District Missed Scan (Last Wk))	2598	5.1055	24.0543	0	609
Indicator6 * (Own Missed Scan - Lowest District Missed Scan (Last Wk))	2598	3.2456	11.1455	0	149

APPENDIX TABLE 2

Correlation Coefficients

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) missedscan	1														
(2) Weekly Sales (000s)	-0.139	1													
(3) Avg Units (per trans)	-0.123	0.175	1												
(4) Emp.Experience (yr)	0.004	0.373	0.022	1											
(5) Supervisor Experience (yr)	-0.032	0.054	-0.044	0.301	1										
(6) % Old Guard	0.078	0.310	-0.017	0.728	0.070	1									
(7) Indicator1: Closest Pilot not District Pilot	0.029	0.017	0.026	0.022	-0.013	0.001	1								
(8) Indicator2: Pilot - Own Missed Scan > 0	-0.284	0.064	0.125	0.084	0.032	0.070	-0.089	1							
(9) Distance to Pilot Store (km)	-0.068	-0.251	0.266	-0.162	-0.040	-0.160	0.380	0.015	1						
(10) Indicator3: Match the Shift of Pilot (Day or Night)	-0.013	0.058	0.069	-0.068	-0.056	-0.052	-0.091	0.057	0.018	1					
(11) Abs(Difference in Population Density of Own - Pilot Store)	-0.029	0.203	-0.029	0.129	0.036	0.111	0.017	-0.022	-0.165	0.066	1				
(12) Difference in Population Density of Own - Pilot Store	0.068	0.107	-0.116	0.127	0.002	0.181	0.113	0.061	-0.103	-0.090	0.004	1			
(13) Own Missed Scan - Lowest District Missed Scan (Last Wk)	0.309	-0.048	-0.033	-0.006	-0.016	0.045	-0.021	-0.263	-0.057	0.046	-0.006	0.029	1		
(14) Indicator4 * (Own Missed Scan - Lowest District Missed Scan (Last Wk))	0.163	0.018	-0.031	-0.009	0.003	0.027	-0.016	-0.155	-0.028	0.038	0.011	-0.006	0.846	1	
(15) Indicator6 * (Own Missed Scan - Lowest District Missed Scan (Last Wk))	0.221	-0.048	-0.022	0.000	-0.038	0.019	-0.012	-0.165	-0.042	0.014	-0.015	0.016	0.303	-0.064	1