Working Paper Series

2017/10/TOM/ACGRE

(Revised version of 2013/114/TOM/ACGRE)

When You Work with a Superman, Will You Also Fly? An Empirical Study of the Impact of Coworkers on Performance

Tom Fangyun Tan
Southern Methodist University, ttan@cox.smu.edu

Serguei Netessine
INSEAD, serguei.netessine@insead.edu

We examine a large operational data set in a casual restaurant setting to study how coworkers' sales ability level affects other workers' sales performance. We find that waiters react non-linearly to their coworkers' ability. In particular, when coworkers' overall sales ability is low, increasing this ability may prompt waiters to redouble both upselling and cross-selling efforts. When overall coworkers' ability is high, however, further increasing their ability may trigger waiters to reduce sales efforts. Our empirical findings imply that to maximize sales, managers should mix waiters with heterogeneous ability levels during the same shift. Through two counterfactual analyses, we find that considering the inverted Ushaped peer effects when optimizing current waiters' schedules without changing their capacity may increase total sales between 1 and 3 per cent at no extra cost.

Keywords: Econometrics; Empirical Study on Staffing; Worker Productivity; Business Analytics; Restaurant Operations; Behavioral Operations Management; Quality/Speed Trade-off

The authors acknowledge financial support from INSEAD and the INSEAD-Wharton Alliance.

When You Work with a Superman, Will You Also Fly? An Empirical Study of the Impact of Coworkers on Performance

Tom Fangyun Tan

Cox Business School, Southern Methodist University, Dallas, Texas, U.S.A.

ttan@cox.smu.edu

Serguei Netessine

INSEAD, Singapore

serguei.netessine@insead.edu

Abstract

We examine a large operational data set in a casual restaurant setting to study how coworkers' sales ability level affects other workers' sales performance. We find that waiters react non-linearly to their coworkers' ability. In particular, when coworkers' overall sales ability is low, increasing this ability may prompt waiters to redouble both upselling and cross-selling efforts. When overall coworkers' ability is high, however, further increasing their ability may trigger waiters to reduce sales efforts. Our empirical findings imply that to maximize sales, managers should mix waiters with heterogeneous ability levels during the same shift. Through two counterfactual analyses, we find that considering the inverted U-shaped peer effects when optimizing current waiters' schedules without changing their capacity may increase total sales between 1 and 3 per cent at no extra cost.

1 Introduction

Although service sectors play a more and more important role in the global economy, they generally suffer from low labor productivity. An OECD study shows that labor productivity in the wholesale and retail trade, hotels and restaurants across OECD countries is typically about three fourths that of industry sectors, such as manufacturing (Freeman, 2008), thus creating opportunities for productivity improvement. One such opportunity lies in effective team management, not only because in service sectors employees often work in interdependent teams/groups[1] (Cohen and Bailey, 1997), but also because good teams often create value

[1] In this paper, we refer to a group of employees simultaneously performing similar tasks together as either a team or a group. Admittedly, the terms 'teams' and 'working groups' sometimes have different definitions in the management literature (e.g., Katzenbach and Smith, 1993), wherein a group is characterized by individual accountability and a team is characterized by both individual and mutual accountabilities. We do not make such a distinction and use the terms interchangeably.

more than the sum of their individual parts. Effective teams tend to promote knowledge sharing, repeated collaborations and creativity (e.g., Argote and Epple, 1990; Uzzi and Spiro, 2005; Guimera et al., 2005), and reinforce individual goals and accountability (Katzenbach and Smith, 1993).

For a long time the operations management literature modeled service employees as having homogeneous (but perhaps random) capacity. Of course, a labor force is typically heterogeneous on the skill/ability dimension, and this consideration has been recently incorporated into labor decisions in settings such as skill-based routing, staffing and scheduling, as well as structural flexibility decisions (see Section 2 for references). Still, most of these studies use analytical models assuming that heterogeneous workers work independently from each other, and little empirical research has been done to understand how to exploit the spillover/peer effects – that is, how team members with heterogeneous skill levels affect each other. Our paper aims to empirically examine these spillovers and incorporate the empirical findings into the design of scheduling/rostering models to understand their implications for operational performance as well as to demonstrate the value of the data analytic approach for labor management. One specific decision we analyze is whether or not to polarize team ability during the same shift and how to schedule the best or worst workers.

Following prior research, we focus on the spillovers between coworkers' ability and the performance of others, which we refer to as peer effects (Chan et al., 2014a). Research has proposed various mechanisms by which spillovers happen, including free riding, competition, monitoring and social pressure. Empirical studies on peer effects in the workplace have typically found linear peer effects, either positive or negative. In this paper, we propose to examine a potential non-linear effect of peers' ability and performance, which may reconcile the seemingly conflicting linear effects found in previous research. We chose to study a large data set of a full-service casual restaurant chain because 1) waiters exhibit wide ability level heterogeneity; 2) waiters have significant influence on sales; and 3) multiple waiters are scheduled to work during the same shift, potentially affecting each other. Specifically, we collect detailed transaction-level data from the restaurant chain's point-of-sales system, which contains approximately 226,350 check-level observations for three restaurants from January 2011 to June 2012. Using an instrumental variable approach, we demonstrate the inverted-U relationship between performance (sales) of the focal employee and coworkers' sales ability and we further show how the labor mix can be leveraged to optimize the employee schedule. Particularly, we show that mixing waiters of various ability levels in the same shift is associated with improved financial performance. Lastly, we conduct two counterfactual analyses of the impact on sales when managers consider the inverted-U peer effects on scheduling: a sales increase between 1 and 3 per cent at no additional cost.

2 Related Literature

Our research contributes mainly to two streams of literature, optimal scheduling/rostering decisions and peer effect studies.

How to schedule workers to meet the stochastic demand belongs to classic labor management problems in services, and we refer our readers to Gans et al. (2003), Akşin et al. (2007) and Van den Bergh et al. (2013) for their excellent literature reviews. Whereas classical models tend to be constrained by assumptions that workers are similar and independent from each other for modeling tractability reasons, recent studies in operations management have developed new approaches to managing a group of heterogeneous workers in various areas, including skill-based routing (e.g., Wallace and Whitt, 2005; Ata and Van Mieghem, 2009; Mandelbaum et al., 2012; Mehrotra et al., 2012; Ward and Armony, 2013); hiring and retention (e.g., Arlotto et al., 2013); structural flexibility (e.g., Hopp et al., 2004; Iravani et al., 2007; Narayanan et al., 2009; Kesavan et al., 2014) and incentive design (e..g, Siemsen et al., 2007; Roels and Su, 2013). Similarly, researchers have also started to incorporate workers' heterogeneity in scheduling/rostering decisions, which is the focus of this paper. For example, Cezik and L'Ecuyer (2008) and Bhulai et al. (2008) devise efficient techniques for scheduling call center agents with different skills and labor costs to handle calls requiring varying skills in order to minimize costs. Bard and Wan (2008) consider workers having non-symmetric movement restrictions among work sites to find the best mix of employees to satisfy demand at minimum cost. Nevertheless, none of these scheduling/rostering papers have considered the spillover effects among the workers. Our paper empirically examines the spillover and then sheds new light on incorporating the empirical findings into designing scheduling/rostering models to demonstrate their implications for operational performance through better scheduling. It is noteworthy that scheduling and staffing decisions are seemingly equivalent but actually different in that the former determines the team composition during a particular shift and the latter decides the number of workers for a shift. For example, Tan and Netessine (2014) study a staffing problem and find an inverted U-shaped relationship between workload and performance. Accordingly, if the workload is less than the optimum, they suggest that reducing the number of workers per hour may not only reduce labor costs but also achieve a sales lift. Unlike that study, this paper focuses on scheduling decisions.

Furthermore, although considerable research has been devoted to making optimal scheduling decisions analytically, less attention has been paid to explicitly evaluating the impact of scheduling on workers' performance. As an exception, in studying retail labor mix decisions, Kesavan et al. (2014) find that increasing temporary labor mix from zero to its optimal value increases sales by 6.78%, and that increasing a part-time labor mix from zero to its optimal value increases sales by 15.04%. However, the focus of their paper is on labor force flexibility, while our paper is about evaluating peer effects within a team. Akşin et al. (2015) find that the new recruits at the London Ambulance Service who have worked with more different partners in the past tend to perform more efficiently during patient pick-up and handover processes than those who have worked with the same partners, because they have benefited from group learning. Thus, their paper primarily studies the diversity of partnership experience, which is again different from the skill heterogeneity and peer effects examined in this paper.

The examination of peer effects has recently attracted attention in the labor economics literature. In a lab setting, Falk and Ichino (2006) discovered that peer effects improved workers' envelope-stuffing productivity as a consequence of peer pressure. In an academic performance setting where students are encouraged to learn from each other, Carrell et al. (2009) posited that the ability of a cohort at the U.S. Military Academy, measured in terms of average SAT verbal score, should have a positive effect on the academic performance of every member of the cohort. In practice, however, after implementing an intervention based on theoretical prescription, Carrell et al. (2013) observed that peer effects turned out to have a negative impact on low-performance students because they tended not to interact with high-performance students. More relevant to our setting is a research on peer effects in the service-oriented workplace by Mas and Moretti (2009) , who study a supermarket register checkout setting, where workers are paid a fixed hourly rate. They find evidence of positive productivity spillovers from highly productive workers because of social pressure. Similarly, Schultz et al. (2010) show that workers on a production line adjust their speed to the average speed of their coworkers since their work stations are interdependent. In a setting without externalities, where workers are paid a piece rate, Bandiera et al. (2010) find that a fruit picker's productivity increases when he/she works with more capable friends, since workers have preferences to socialize with their friends. Chan et al. (2014a) analyze the sales performance of the salespeople at cosmetics counters and argue that the incentive scheme determines the direction of peer effects. According to their findings, while team-based commissions produce positive peer effects because workers may help each other, individual-based commissions create negative peer effects because strong salespeople may gain customers from lower ability coworkers. Even in a knowledge-based workplace, Staats et al. (2015) find that cardiologists are more likely to choose the same treatment procedures as more experienced colleagues because the experienced surgeons exert group pressure. Despite these seemingly conflicting linear peer effects, there is some evidence that the linear peer effects may not even exist. For example, Guryan et al. (2009) analyze whether playing partner's ability affects performance among PGA golfers and find the estimates statistically insignificant.

Our study contributes to this stream of literature in two significant ways. First, most peer effect studies tend to focus on linear peer effects, the results being positive, negative or silent, whereas our study focuses on non-linear peer effects, which may reconcile the conflicting findings noted above. Second, although considerable research has been devoted to assessing peer effects, much less attention has been paid to the implication of peer effects for labor decisions in service operations, or the value of incorporating peer effects into labor decisions. For example, in Mas and Moretti (2009), the retail managers are not responsible for assigning individual workers to particular shifts. To bridge the gap between empirical peer effects studies and the optimal labor decision literature in service operations, we build quite realistic scheduling/rostering models that incorporate peer effects to make optimal scheduling decisions.

3 Hypotheses Development

One in three Americans have worked in the restaurant industry at some point in their life (Mill, 2006), hence restaurant waiters come from different backgrounds, and display a wide heterogeneity in ability. In this paper we focus on waiters' sales skill levels because 1) sales have a direct impact on both restaurants' and waiters' income; 2) any increase in sales is particularly significant in the casual dining industry, where profit margins are only 3% to 9%; 3) a well-executed sales job will substantially enhance customers' dining experience. A high-ability waiter tends to have a pleasant personality and attitude. He/she has a thorough knowledge of both the food and the wine, and thus the table-side confidence to successfully make suggestive sales. A high-ability waiter is able to "read" diners and anticipate their needs. For example, he/she rarely leaves diners' glasses empty, maximizing the opportunities to sell beverages and wines. By contrast, a low-ability waiter may simply mention the "$9.99 special" once diners are seated, failing to sell more expensive items or remind them about ordering extra items.

Waiters with heterogeneous sales skills often work together during a shift. Although sometimes considered "independent business people" (Walker, 2007) because the majority of their income comes as tips from the tables they serve, they are also trained to collectively contribute to the restaurant as a whole. In other words, team interaction can influence an individual's service performance. Effective teamwork is important for restaurant operations since it improves customer satisfaction, encouraging repeat visits and therefore increasing long-term financial performance; it promotes a sense of achievement, equity and camaraderie, keeping waiters motivated and satisfied and thus reducing turnover, which can be costly; it facilitates learning in the workplace, strengthening collaboration. Given the importance of teamwork, waiters are often referred to as "team members" in a restaurant.

Desirable as high ability is, it does not always precisely translate into higher performance, which also requires waiters' contextual efforts. Similar to many other multitasking agents (e.g., DeHoratius and Raman, 2007), waiters in a casual restaurant expend two important types of sales effort to serve seated diners, which consist of both upselling more expensive items and cross-selling additional items, which both lead to higher sales amounts per check. Although such efforts are not directly observable, they may be inferred from observable performance metrics (i.e., the sales of each meal), for which we develop hypotheses about the peer effects below. We categorize the theories into two main types: *positive spillover* effects and *negative spillover* effects.

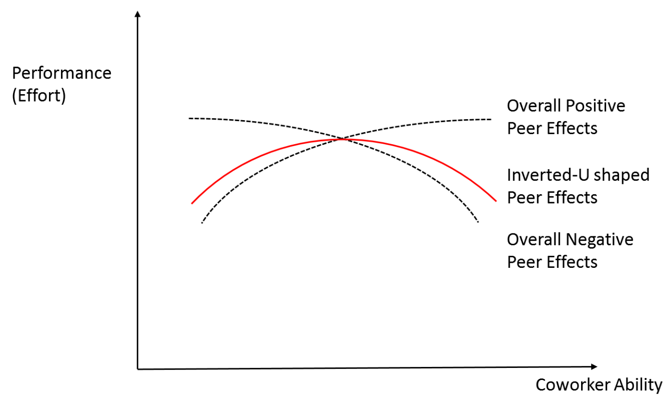
Positive Spillover Effects Positive spillovers are defined as the phenomenon that high-ability workers improve the performance of their coworkers (Mas and Moretti, 2009). Behind this phenomenon there are at least three theories from economics and social psychology. First, due to social pressure, a worker may experience disutility if s/he is observed behaving noncooperatively by their peers for fear of sanctions, shame or reputational concerns (Mas and Moretti, 2009). This disutility is likely to intensify as coworkers become more and more competitive (of higher ability), which will overshadow the focal worker and make him/her perceived as even more noncooperative. In order to reduce this disutility, a worker may expend more effort to catch up with those higher-ability coworkers. Consequently, social pressure may help mitigate the free-riding problem, especially when employees work as a team (Kandel and Lazear, 1992). Second, knowledge spillover implies that information about how to do a job well may transfer from one worker to the next (Argote and Ingram, 2000; Moretti, 2004a,b; Chan et al., 2014b). This advantageous knowledge is usually possessed by the high-ability workers, who may choose to share it with lower-ability coworkers for prosocial/altruistic reasons (Itoh, 1991, 1993; Siemsen et al., 2007), or it can be learned by low-ability workers through observation (Song et al., 2015). Third, social comparison suggests that people compare themselves to others for self-evaluation (Festinger, 1954). When comparing with high-ability peers, people may exhibit behind-averse behavior by working harder to minimize the disparity (Roels and Su, 2013; Kuziemko et al., 2014). For example, workers may seek out new ways of working and deliberately learn from high-ability coworkers (Pisano et al., 2001; Nembhard and Tucker, 2011). Note that this drive to work harder does not necessarily stem from being perceived as noncooperative, which is different from the social pressure theory. Positive spillovers could happen to waiters who work as a group during the same shift. Although they may be considered independent workers because they earn their tips only from the tables that they serve individually, waiters are trained to collectively contribute to the whole restaurant (Walker, 2007). If a waiter particularly lags, other waiters may report him/her to management or ostracize him/her socially, creating social pressure for that waiter to expend more effort. In addition, waiters may learn from the higher-performing coworkers either by watching or by exchanging ideas during the service meetings before the shift. Furthermore, waiters compare their tips at the end of the shift, which should motivate behind-averse waiters to improve their performance.

Negative Spillover Effects Unlike positive spillover effects, negative spillover effects suggest high-ability workers reduce the performance of their coworkers (or low-ability workers increase the performance of their coworkers). We suggest three theories that may predict such effects. First, a free rider problem may appear in the teamwork setting when workers are serving the same pool of customers. Focal workers are more likely to be triggered to reduce their effort when working with high-ability coworkers than with low-ability ones because they may think their high-ability coworkers will serve more customers and generate most of the sales (Hölmstrom, 1979; Chan, 2016). Second, when comparing with low-ability coworkers, people may exhibit ahead-seeking behavior by working harder to maintain their top position (Brown et al., 2007; Roels and Su, 2013; Kuziemko et al., 2014). Third, overly capable coworkers may provoke anti-productive emotions among other workers. When coworkers are excessively capable, they may pose a threat that hinders workers from reaching their goals, which may further reduce their motivation and commitment (e.g., O'Connor et al., 1984; Barankay, 2012). Comparison with highly capable coworkers may also create negative feelings about oneself (Buunk et al., 1990) and unhappiness when those highly capable coworkers have higher earnings (Luttmer, 2004), all of which may lower motivation and reduce effort. Instead of working hard, demoralized workers may try to sabotage the high-ability coworkers (Lazear, 1989; Chen, 2003), negatively impacting their performance. When waiters work with other highly capable coworkers, they may similarly have the aforementioned free riding problem and feel anti-productive emotions. They may become demoralized because the highly capable coworkers hinder them from reaching their desired goals, which may include getting assigned to favorable table sections or winning the explicit/implicit sales contest during the shift. Workers may also feel disappointed about themselves when benchmarking against extremely capable coworkers. Consequently, they may give up devoting more effort to service. On the flip side, when working with low-ability coworkers, those waiters may have extra incentive to work harder to pull ahead because restaurant managers often run (unofficial) sales tournaments to reward the waiters selling the most during the day.

Prediction Regarding Overall Impact of Peer Effects on Sales Figure 1 shows the combination of overall positive and negative spillovers in creating our hypothesis of the peer effects on sales performance. First, both the positive and negative spillover effects should be concave in coworkers' ability because of diminishing returns to scale. For example, a waiter can indeed learn more from a more capable coworker, but the rate of improvement should decrease because it becomes more difficult to learn new skills as one's performance improves (i.e., a ceiling effect). Furthermore, when coworkers' overall sales ability is low, positive spillovers may dominate negative ones because 1) low-ability coworkers cannot sell enough to let focal workers free ride without being noticed; 2) very low ability coworkers may not create enough drive for focal workers to seek ahead (it does not cost much effort to be the best); 3) low-ability coworkers may not generate enough anti-productive emotions for the focal worker. For these same reasons, when the overall coworkers' sales ability is high, negative spillovers may dominate the positive spillovers. Therefore, we propose:

HYPOTHESIS 1 (H1): *As coworkers' sales ability increases, the sales performance of the focal employee will first increase and then decrease: that is, there is an inverted U-shaped relationship between coworkers' ability and sales.*

Figure 1: Combining Overall Positive and Negative Spillovers



Prediction Regarding the Effect of Coworker Proximity on Peer Effects Positive and negative spillover effects require that the focal worker can observe his/her coworkers and be observed by them. If the worker can neither observe nor be observed by other coworkers, he/she will hardly feel the social pressure from the high-ability coworkers because the fear of sanctions or shame do not reach him/her. Similarly, both social comparison and anti-productive emotions are not likely to materialize if the worker simply does not know the ability level of his/her coworkers. Research further suggests that proximity facilitates communication, coordination, mutual support, effort and cohesion in the team (Hoegl and Proserpio, 2004). Examining U.S.-based pharmaceutical plants over a 13-year period, Gray et al. (2015) find that physical proximity through geographical collocation between manufacturing and R&D activities improves conformance quality. In addition, in a study of supermarket cashiers, Mas and Moretti (2009) find that only those cashiers who can directly observe fast coworkers in front of them may experience an improvement in their productivity because of social pressure.

On the other hand, knowledge spillover effects do not necessarily rely on the observability condition because high-ability workers can still share their knowledge at work. In the restaurant setting, waiters are typically assigned to specific table sections, some of which are closer to or farther away from each other. Although waiters who are in close proximity to each other are more likely to observe each other, they are still aware of the entire team of the shift because they meet them during the pre-shift meetings and near the kitchen and the common area. Hence, we do not suggest that the coworkers whose table sections are far away exert no peer effects at all on the focal worker. Instead, we hypothesize that distance weakens the peer effects:

HYPOTHESIS 2 (H2): *Coworkers whose work sections are in proximity to the focal employee have stronger peer effects than those farther way.*

Prediction Regarding the Effect of Team Heterogeneity on Sales Our last prediction concerns the team composition. Assuming the inverted U-shaped peer effects as hypothesized in H1, polarizing the team ability (i.e., rostering all the high-ability or low-ability workers in a shift, separately) may cause it to be either too high or too low, the two suboptimal areas of the inverted-U curve. Admittedly, an all-mid-range team is indifferentiable from a high-low mix team in terms of the average team ability. However, in practice, always forming an all-mid-range team across all shifts is infeasible because a manager must roster from only a fixed pool of heterogeneous workers. We therefore need to understand the implication of heterogeneity for the team ability of all shifts under worker capacity constraints. To explain the intuition, we use a hypothetical example. Suppose we have two workers of each of the three skill levels (high, medium and low). Roster decision 1 has the following combinations: shift 1: low-low; shift 2: medium-medium; shift 3: high-high. By contrast, roster decision 2 has the following combinations: shift 1: low-high; shift 2: low-high; shift 3: medium-medium. We argue that roster 2 has a better performance than roster 1 because the average team ability in roster 2 is more likely to be in the middle than the average team ability in roster 1. In addition, roster 2 has a higher average team heterogeneity than roster 1. To summarize, average heterogeneity per shift is generally associated with higher performance, although the heterogeneity of each shift in the preferred roster is not necessarily always higher than that of every shift in another roster, as can be seen in the medium-medium combination.

Besides the inverted U-shaped peer effects argument, some previous studies report positive effects of team heterogeneity on performance. For example, even adjusting for the average team ability, Hamilton et al. (2003) find that more heterogeneous teams in terms of ability were more productive in manufacturing garments because of mutual learning and intrateam bargaining. Similarly, Shafer et al. (2001) show that worker heterogeneity in terms of learning rates produces higher output among a group of workers operating independently of one another. For these reasons, we posit:

HYPOTHESIS 3 (H3): *Worker heterogeneity in a team increases sales performance.*

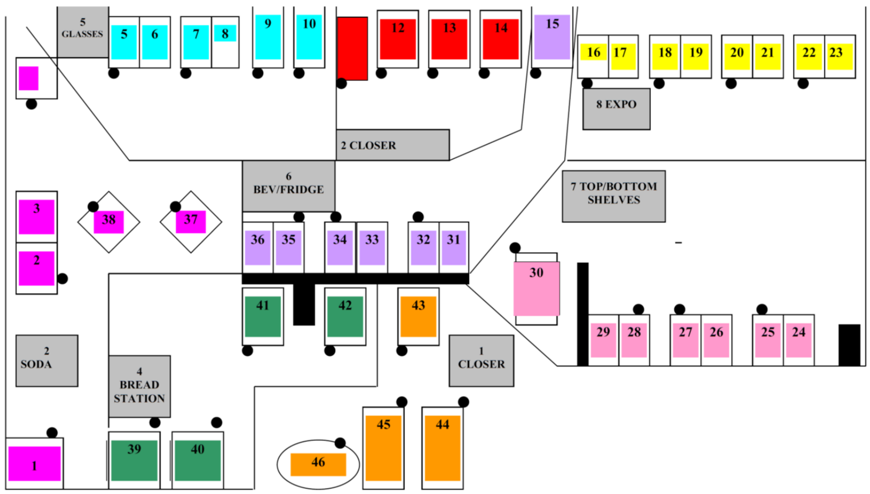
4 Empirical Setting, Variables and Descriptive Statistics

4.1 Empirical Setting

Similar to other family-style restaurants such as TGI Friday's and Applebee's, our empirical setting is a restaurant chain that offers casual American-style dining with table service, located in the Boston suburbs. In this setting, waiters are responsible for waiting on assigned tables, from which they earn gratuities in addition to a restaurant minimum hourly wage. Waiters are typically scheduled to work on shifts of varying lengths (e.g., four hours). Opening hours of the restaurants are from 11:30am to 10:00pm, Monday to Thursday, and from 11:30am to 11:00pm, Friday to Sunday. From three restaurants of this chain we collected 18 months of point-of-sale (POS) data from January 2011 to June 2012, including detailed information about waiters, sales, party size, and service start and end time for each check. The floor plan of one restaurant is shown in Figure 2, where tables are numbered and sections are color coded. As indicated by the black dots, each waiter is typically assigned up to four parties in a section. We were able to obtain the floor plan of only this particular restaurant because we worked with this restaurant to implement a new scheduling system (more details on this system are provided in Subsection 5.2).

Since our empirical analysis focuses on the main dining room, where peer observation and interaction are most likely to happen, we exclude the bar and take-out services. Further, we drop the transactions which include the day's top and bottom 7.5% of checks to reduce the influence of outliers (e.g., very large parties and private events). Our final data set is comprised of approximately 226,350 check-level observations. This setting has a number of advantages for examining peer effects on individual work performance. First, waiters have heterogeneous abilities, creating a unique challenge and opportunity. Second, we know who is

Figure 2: Typical Floor Plan



working at any moment in time, so that we can identify the working group and the peers to further assess the implications for labor decisions.

4.2 Variables and Summary Statistics

Our main analysis is conducted at the individual check level because this granularity of analysis contains more information about waiters' behavior (cross-selling and upselling) in handling each dining party. Hence, we provide check-level variable definitions in this subsection. We use sales to measure waiters' performance because it is evaluated by the management and captures the dining aspects over which waiters can exert leverage. We conduct additional analysis at the hour level in order to check the robustness of the check-level results and to generate implications for the total performance from a restaurant perspective.

$Sales_i$. Our dependent variable measures the sales (in dollars) of check *i,* which is exclusively assigned to one waiter in our focal restaurant. We consider it to reflect service quality, which is concerned less with service accessibility as often assumed in the service operations literature than with a standard for service contents.

$CoworkerAbility_i$. Similar to previous literature (e.g., Mas and Moretti, 2009), we construct the key independent variable $CoworkerAbility_i$ in two steps. For the first step, we employ a fixed-effects model to estimate the intrinsic sales ability of waiter j in a given month m, $OwnAbility_{jm} = \theta_{jm}$ (θ_{jm} can also be negative). Waiters' sales abilities may fluctuate over time for reasons such as learning (e.g., Argote and Epple, 1990; Lapré et al., 2000), forgetting (e.g., Shafer et al., 2001), and task variation (e.g, Wiersma, 2007; Staats and Gino, 2012). Since the focus of our study is peer effects instead of learning, forgetting or task variation effects, we do not separately identify those factors of intrinsic abilities. Our approach is to control for all these effects through estimating waiter *j's* intrinsic sales ability during each month, which reflects both a relatively stable underlying ability within a month and a variable ability over a longer time period.

In particular, we divide our data into 18 months (from January 2011 to June 2012). Then following Chan et al. (2014a), who measure a focal salesperson's permanent sales productivity (i.e., dollar sales per hour), we specify the following fixed-effects model to estimate the intrinsic sales ability (i.e., waiter *j's* average sales value in a check) and run this fixed-effect model for each month m, separately:

$$\sum_{i \in jtm} \frac{Sales_i}{PartySize_i} = \beta_0 + \theta_{jm} + \beta_1 Controls_i + \varepsilon_{jtm} \quad \forall m = 1, \cdots, 18. \tag{1}$$

In the specification, $\sum_{i \in jtm} \frac{Sales_i}{PartySize_i}$ is measured by averaging the sales of all the checks that are opened during hour t and handled by waiter j over all the diners who contribute to these sales (i.e., the waiter*'s* per person average dollar sales (PPA) in each hour). PPA is a financial measure often used by restaurant management (Mill, 2006). We calculate PPA in dollars instead of log-transforming it for interpretation purposes. We estimate the intrinsic sales ability at the hour level because 1) waiters typically work hourly shifts; 2) hourly aggregation instead of daily aggregation ensures both an adequate sample size in each hour for each waiter and enough observations over time for statistically significant estimates. In addition, $Controls_i$ include $DayWeek_i$, $Hour_i$, $YearWeek_i$ and $Store_i$ to adjust for the time, date and location factors (their definitions are presented in the Control Variable paragraph of this section), such as demand and coworker composition. We include store-specific time-invariant factors to control for unobserved heterogeneity among stores, such as the income level of the neighborhood and other time-invariant omitted variables. Ideally we would like to explicitly control for all the combinations of coworker composition, as in Mas and Moretti (2009). However, we encounter the difficulty that there are hundreds of unique worker combinations in our data and many combinations have only limited repeat observations, which will inflate standard errors and thus affect the estimation of peer effects. Nevertheless, $Controls_i$ and the time-varying ε_{jtm} should implicitly adjust for the influence of coworkers. The workers in our empirical setting are scheduled to work on an hourly basis. so each combination of $DayWeek_i$, $Hour_i$, $YearWeek_i$ and $Store_i$ therefore uniquely identifies each worker combination, which is similar to the problem of estimating individual fixed effects and firm fixed effects in longitudinal data containing unique firm and worker identifier (Abowd et al., 1999).

For the second step, we take the average of the intrinsic sales abilities of the coworkers working during the same hour as waiter *j*, who opens the focal check *i* being analyzed, to form a peer effect variable, $CoworkerAbility_i$. In other words, $CoworkerAbility_i = \bar{\theta}_{-jm} = (1/n)\sum_{k\neq j}\theta_{km}$, where *n* is the number of coworkers. For example, suppose check *i* is handled by waiter *j*. When the check is opened, waiter *j* has four coworkers, whose intrinsic sales abilities are \$1, \$2, -\$2, and \$4, respectively. Our peer effect measure $CoworkerAbility_i$ is (\$1 + \$2 - \$2 + \$4)/4 = \$1.25. Averaging coworkers' sales abilities reflects the absolute effect of the peers' sales abilities, which is consistent with prior work by Mas and Moretti (2009) and Carrell et al. (2009). These absolute peer effects are comparable across the three restaurants because the three stores belong to the same chain offering standardized menus. Equally important, we consider an alternative peer effects measure, which represents coworkers' ability relative to the focal waiter's. In other words, we create an alternative independent variable, $RelativeAbility_i = CoworkerAbility_i - OwnAbility_i$, which is also used in previous peer effects literature (e.g., Schultz et al., 2010; Chan et al., 2014a).

Note that we measure the peer effects in terms of coworkers' intrinsic sales abilities. An alternative and equally interesting independent variable would be coworkers' contemporaneous performance[2] (i.e., performance during the same shift rather than some average performance). In this study, we use the dependence of the peer effect on waiters' intrinsic sales ability because 1) waiters talk to each other and often compare their

[2]Similar to Mas and Moretti (2009), we use 'effort' and 'performance' interchangeably in this paper.

tips at the end of the shift, so they may have no knowledge about who are high-performing waiters currently but they may know who is high performing on average; 2) waiters may be so busy waiting their own tables that they have little time to observe coworkers' contemporaneous performance; 3) performance shocks that affect all waiters at a particular moment can create a spurious relationship between focal waiter and coworker performance; 4) we replaced the monthly ability measure with average hourly per person average sales (a contemporaneous performance measure) to construct the peer effects, only to find the coefficients were statistically insignificant. Nevertheless, we do not argue that the contemporaneous performance has absolutely no effect on the focal worker's performance because it is correlated with long term ability. Rather, we interpret the results as an evidence that the *presence* of certain coworkers (e.g., superstars) has a stronger effect on the focal worker's performance than their *performance*. Most importantly, only intrinsic sales ability can be used in proactively scheduling waiters because it can be calculated in advance.

Control Variables. We control for several variables that, according to previous literature, could affect waiters' performance. Variable $OwnAbility_i$ is θ_{jm} estimated from Model 1, that is, the intrinsic sales ability of waiter j responsible for check i. When we include this variable, we intrinsically consider the focal waiter's ability even with the absolute peer effects measure. Variable $PartySize_i$ controls for the number of diners in a particular party i, which should positively affect sales. Variable $AbilityStDev_i$ is the standard deviation of the sales abilities of all the waiters working when check i is opened. We use this variable to adjust for ability dispersion/heterogeneity, which is known to affect worker performance in Chan et al. (2014a). We do not use the coefficient of variation to measure the dispersion because our intrinsic sales ability has both negative and positive values. In addition, we control for a one-hour lagged effect of $CoworkerAbility_i$, calling it $LagCoworkerAbility_i$, because the peer effect may propagate over time (Mas and Moretti, 2009; Carrell et al., 2009). Furthermore, following Tan and Netessine (2014), who find a non-linear relationship between waiters' workload and their performance, we control for the individual workload $AvgTables_i$ and its quadratic form. Variable $AvgTables_i$ is the average number of tables (parties) that a waiter handles simultaneously with the focal check i being analyzed. For instance, suppose check i lasts 50 minutes, when it shares its waiter with another table (party) for 10 minutes. The workload measure $AvgTables_i$ is (50 min + 10 min)/(50 min) = 1.2 tables. Calculated in the same way we calculated the individual workload, variable $StoreTables_i$ is the average number of tables occupied during check i in the entire restaurant, which is used to control for the storewide traffic/congestion. This variable also controls for the potential mechanistic peer effects when high-ability waiters disproportionately increase the load on the kitchen and affect other waiters. Finally, we include additional fixed effects of the time/date/location of check i to control for temporal and spatial factors, such as demand. In particular, we include a categorical variable $Hour_i$ (11am, 12pm, ..., 11pm), the hour when check i was opened, to control for systematic intra-day differences in demand. We include another categorical control, $DayWeek_i$, indicating the day of the week (Sunday, Monday, ..., Saturday) because weekends are usually busier than weekdays. In addition, in order to adjust for seasonality and economic trends, we use a categorical control variable, $YearWeek_i$, which starts at one from the first week of January 2011 and ends at 79 in the last week of June 2012. Finally, we include a categorical variable $Store_i$ for each store i to control for time-invariant aspects of store fixed effects (e.g., location, traffic).

4.3 Descriptive Statistics

Table 1 shows the descriptive statistics of the check-level variables. Each check has an average sales of *Sales*=\$45. There are on average $PartySize = 2.51$ diners in each check, which translates to \$45/2.51 $\approx$ \$18 per diner. Furthermore, there is a considerable heterogeneity in coworkers' sales ability and focal waiters' intrinsic sales ability. For example, the coworkers' sales ability ranges from *CoworkerAbility*=-\$4.64 to \$8.09, with the focal waiters' intrinsic sales ability ranging from *OwnAbility*=-\$13.59 to \$15.8. Each waiter on average handles $AvgTables = 2.32$ tables simultaneously, and the entire store has on average $StoreTables = 16.36$ tables occupied (store capacity is approximately 40 tables). Note that *Sales* seems to be right skewed (mean = 45 > median = 39), suggesting that the residuals are asymmetrically distributed. We therefore transform *Sales* into its natural logarithm, a commonly used technique (Velleman and Hoaglin, 1981; Albright and Winston, 2014), to make the residuals more symmetrically distributed to form a bell shape, because natural log transformation can squeeze the large values of the dependent variables together and spread the small values out. In other words, log transformation increases the normality of the errors, which ensures that the model inference is valid. In addition, natural log transformation makes it easier to interpret the monetary effect of *Sales* in terms of percentage changes. As a robustness check, we attempted the same analysis without this transformation with similar results.

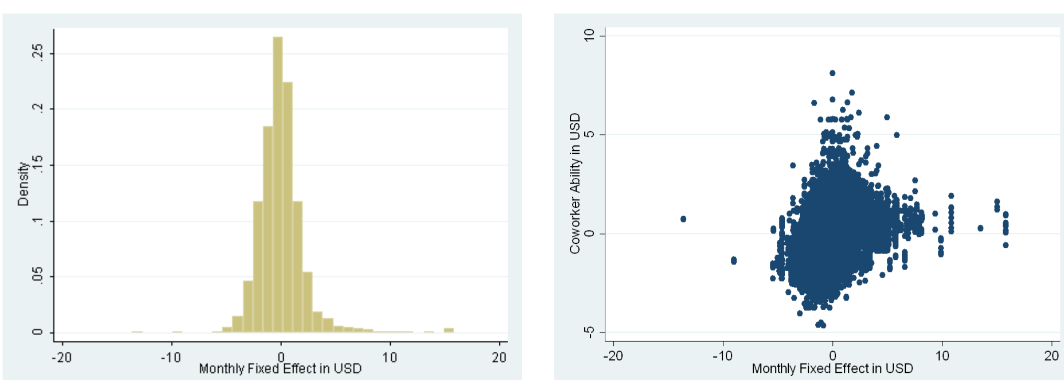
Table 1: Summary Statistics of Check-Level Variables

	Sales	*Coworker Ability*	*OwnAbility*	*PartySize*	*Ability StDev*	*LagCoworker Ability*	*AvgTables*	*StoreTables*
N	220,923	220,923	220,923	220,923	220,923	206,257	220,923	220,923
Mean	45.01	0.02	0.03	2.51	1.13	0.03	2.32	16.36
Stdev	22.87	1.02	1.50	1.04	0.51	1.03	0.80	6.48
Min	5.11	-4.64	-13.59	1.00	0.00*	-3.51	1	1
P5	18.27	-1.73	-2.32	1.00	0.47	-1.74	1	5.04
P50	39.67	0.31	0.01	2.00	1.06	0.32	2.27	16.94
P95	91.22	1.31	2.42	5.00	2.01	1.31	3.72	26
Max	149.97	8.09	15.80	8.00	6.02	6.59	10.58	37.32

*The exact value is 0.00039.

Figure 4a shows the histogram of the sales ability distribution, illustrating a wide variation in waiters' intrinsic sales abilities. Figure 4b further displays the scatter plot of waiters' own intrinsic sales ability (*OwnAbility*) and their coworkers' average sales ability (*CoworkerAbility*). There seems to be no apparent relationship between *OwnAbility* and *CoworkerAbility*. Hence, we need a rigorous econometric approach to examining their relationship, while controlling for other factors.

Figure 3



(a) Histogram of Intrinsic Sales Ability (b) Scatter Plot of Own and Coworker Sales Ability

Table 2 shows the correlations of the check-level variables. As expected, log(*Sales*) is positively associated with *PartySize* (correlation = 0.5985). It is noteworthy that the correlation between the dependent variable and *CoworkerAbility* is low (-0.0046). However, the low correlation does not necessarily indicate that peer effects have no relationship with the focal waiter's performance because their relationship may be non-linear. In addition, the correlations between other predictors are generally quite low, except for the correlation between *LagCoworkerAbility* and *CoworkerAbility,* which is quite high (0.8897), but we compute the variance inflation factors to find them to be below 10, indicating that we are not likely to have multicollinearity problems.

Table 2: Correlation Matrix of Check-Level Variables

	log (*Sales*)	*Coworker Ability*	*OwnAbility*	*PartySize*	*AbilityStDev*	*LagCoworker Ability*	*AvgTables*
log (*Sales*)	1.0000						
CoworkerAbility	-0.0046*	1.0000					
OwnAbility	0.0269*	0.4961*	1.0000				
PartySize	0.5985*	-0.0719*	-0.0990*	1.0000			
AbilityStDev	0.0531*	0.2957*	0.1887*	0.0058*	1.0000		
LagCoworkerAbility	0.0106*	0.8897*	0.6044*	-0.0758*	0.2647*		
AvgTables	-0.0100*	0.1220*	0.0939*	-0.0418*	0.0365*	0.1187*	
StoreTables	0.1733*	0.1689*	0.1078*	0.0843*	0.1290*	0.1731*	0.3440*

*Significant at the 0.05 level.

5 Empirical Analysis Strategy and Results

5.1 Performance Analysis

Unlike previous studies that analyze the data either at hourly or other aggregate level (e.g., Mas and Moretti, 2009; Chan et al., 2014a), we conduct our main analysis at the check level because granular-level analysis tends to be more informative than aggregate-level analysis and we have sufficient data. In particular, we employ the following ordinary least squares (OLS) regression models:

$$\begin{aligned}\log(Sales_i) \quad = \quad & \alpha_0 + \alpha_1 CoworkerAbility_i + \alpha_2 CoworkerAbility_i^2 + \alpha_3 OwnAbility_i + \\ & \alpha_4 PartySize_i + \alpha_5 AbilityStDev_i + \alpha_6 LagTeamAbility_i + \\ & \alpha_7 AvgTables_i + \alpha_8 AvgTables_i^2 + \alpha_9 StoreTables_i + \alpha_{10} Controls_i + \varepsilon_i. \end{aligned} \tag{2}$$

The independent variables *CoworkerAbility* and *CoworkerAbility*^{2} are centered around their means for interpretation purposes. The coefficient of $CoworkerAbility_i^2$ (e.g., α_2 in the sales model) will be negative if there is an inverted U-shaped relationship between peer effect and sales. In addition, the critical point of the performance measure is expected to be at $-\alpha_1/(2\alpha_2)$. In addition, we replace *CoworkerAbility* and its quadratic term with the alternative peer effects measure, namely *RelativeAbility* and its quadratic term, respectively. $Controls_i$ represents the same control variables as in Model 1. We also calculate heteroscedasticity-consistent standard errors so as to allow the fitting of our model to contain potential heteroscedastic residuals.

Although useful as a preliminary estimator, this regression model may not address potential omitted variable bias towards sales estimation. The potential omitted variables should affect both the performance measure and the team composition decisions. In other words, those omitted variables, such as consumers' price sensitivity or their intrinsic level of hunger, will not bias our estimation because this type of consumption-behavior-related factor is likely to be uncorrelated with team composition. Rather, we highlight one significant omitted variable that is related to team composition scheduling. Managers' unobserved demand forecast should be positively correlated with sales. In addition, it should affect team ability, but the direction of the correlation may be ambiguous ex ante. On one hand, managers may be inclined to schedule waiters with high sales ability to work during high-sales days either to match the demand or to reward high-performing waiters. On the other hand, managers may wish to schedule waiters with high sales ability to work during low-sales days because the extra sales improvement for high-performing waiters may be more significant during low-sales shifts than during high-sales shifts. Hence, the direction of the omitted variable bias in the

OLS sales model is inconclusive a priori. Although the direction of these biases cannot be determined ex ante, the omitted variable may still cause inaccurate estimations, so we performed additional Hausman endogeneity tests and rejected the null hypotheses that those peer effect measures were exogenous. In order to alleviate these biases, we turn to an instrumental variable two-stage-least-square (2SLS) approach (Angrist and Krueger, 1994) in the next subsection.

5.2 Instrumental Variable 2SLS Estimation

We rely on an instrumental variable 2SLS approach to address the endogeneity issues because it can provide consistent estimates of the dependent variables using a large sample (Angrist and Krueger, 1994). For an instrumental variable to be valid it should satisfy both relevance and exclusion restriction assumptions (Wooldridge, 2002), which means that it should be uncorrelated with the error (i.e., exclusion restriction) and correlated with the endogenous regressor (i.e., relevance). We introduce two types of instruments, which should satisfy these two conditions. First, we observe an exogenous shock to the scheduling decision during our study period. In the middle of 2011, one of the three restaurants switched to a computer-based scheduling system instead of relying on managers' discretionary decisions. The new computer-based system does not explicitly advise which waiters should be scheduled for each shift; however, it analyzes 13-week historical sales data to forecast the demand for waiters for the next week. The new computer-based staffing level forecast is likely to be different from a manager's forecast because it analyzes more data and tends to be more consistent. Because of the adjusted staffing level, the staffing capacity and team composition may mechanistically change accordingly. For example, the team composition of an eight-waiter team would be, by definition, different from a nine-waiter team. In addition, the ability distribution may change over time because of changing staffing capacity. Consequently, the new scheduling system should affect the average team ability and coworker ability, thus satisfying the relevance condition. Of course, we do not know ex ante whether the scheduling system will increase or decrease the average team ability and will rely on the first stage regression to show its ex post effect. Furthermore, we expect the implementation of the new system to affect sales only through team-composition decisions because diners do not observe the implementation of this labor scheduling system, hence sales should not be directly affected. Therefore, the implementation of the system should also meet the exclusion restriction condition. To operationalize the instrument, we create a dummy variable, *System*, which equals one for all the checks affected by the new scheduling system, and zero for all other observations.

We supplement our analysis using another type of instrumental variable, the lagged values of the endogenous independent variables (e.g., Bloom and Van Reenen, 2007; Siebert and Zubanov, 2010), In particular, we first construct the average team ability during the same hour t as check i opened, $TeamAbility_{it} = 1/n \sum_{j \in t} OwnAbility_j$, where j is one of the n waiters who worked during hour t in the same restaurant. Then *LWTeamAbility* and $LWTeamAbility^2$ are computed to represent the *TeamAbility* and $TeamAbility^2$ of the same restaurant during the same hour of the previous week to be used as instruments for the current week.[3] As an illustration, suppose check i was opened at 7:30pm on 8/8/2011 at restaurant k. Its instrument is *TeamAbility* of the 7:00pm slot on 8/1/2010 at restaurant k. We also mean-center these instruments for interpretation purposes. We elect to use the one-week lag because the focal restaurants schedule workers one week in advance and these weekly schedules tend to be quite stable (the correlation between *LWTeamAbility* and *TeamAbility* is about 0.8). For this reason, the weekly lagged variables should correlate with the current team composition and therefore *CoworkerAbility*, satisfying the relevance condition. Since the scheduling decisions from a week ago should not determine the unobserved factors for sales during the current week, these lagged instrumental variables should also satisfy the exclusion restriction condition. It is true that the lagged team composition may not be ideal in the event of common demand shocks that are correlated over time. We adjust for these common demand shocks, which are basically trends (Villas-Boas and Winer, 1999), in our models with the categorical control variable *YearWeek*, thus lessening this concern. Additional relevant statistics and further discussion to show the validity of these instruments are provided after the main results are presented in Subsection 5.3.

[3]Note that we are unable to construct a lagged *CoworkerAbility* variable because a waiter does not always work the same shift every week.

5.3 Results

Table 3 shows the results of our check-level analysis of the impact of coworkers' ability on waiters' performance using two measures of peer effects. In all the models, the coefficients of the quadratic terms (i.e., *CoworkerAbility*^{2} and *RelativeAbility*^{2}) are consistently significant and negative in both OLS and 2SLS estimations (-0.0023, -0.0075, -0.0014, and -0.0274, respectively), providing support for H1, which states that peer effect has an inverted U-shaped relationship with sales. As a robustness check, we test the model specification by including only the linear terms *CoworkerAbility* and *RelativeAbility,* separately, with control variables. The coefficients turn out to be insignificant in both OLS and 2SLS models, further suggesting a non-linear relationship between coworkers' ability and waiters' sales performance.

In addition, the linear term *CoworkerAbility* is statistically insignificant in the OLS model, but it becomes significant and positive (0.0217) after we correct the potential endogeneity bias using 2SLS estimation. Interpreting this coefficient and the coefficient of its quadratic term, we find that the critical average coworker ability is about $0.0217/(2 \times 0.0075) \approx \$1.4$, which is close to one and a half standard deviations ($1.021) above the sample mean ($0.02). We further calculate that changing the current average coworkers' ability to the optimal value would have generated ($0.0217 \times 1.4 - 0.0075 \times 1.4^2 \approx 1.6\%$) sales lift per check for the focal waiter on average, holding party size and other factors constant. By contrast, the linear term *RelativeAbility* is statistically insignificant in both OLS and 2SLS models, suggesting that the critical point is right at the sample mean.

We believe that the absolute measure is more appropriate in our setting than the relative measure (e.g., Chan et al., 2014a[4]) for the following reasons. First, the adjusted-R^2 is 0.438 in the absolute measure model, while it is 0.398 in the relative measure model, suggesting that the former model has a better goodness-of-fit. Second, waiters may not have an accurate evaluation of their own sales ability for such reasons as superiority bias (Hoorens, 1993), meaning that humans tend to overestimate their own ability and discount others' abilities, thus underbiasing the relative peer effects. Third, our absolute measure model controls

[4]The relative measure is probably more appropriate in the setting of Chan et al. (2014a) because they compare cross-counter peer effects.

for the focal worker's ability through *OwnAbility,* taking the focal worker into consideration, which means it contains the same information the relative measure model. Fourth, as a robustness check, we add an interaction term of the focal worker's ability to the peer effects and find that the $CoworkerAbility^2$ and $RelativeAbility^2$ are both significant and the interaction terms are insignificant, suggesting that the inverted U-shaped peer effects should be applicable to any worker on average. For these reasons, we choose to analyze the absolute measure throughout the rest of the paper and we use the results of the relative peer effects measure here as a robustness check.

For the control variables in the sales models, as expected, *OwnAbility* is positively associated with sales. In particular, its coefficient in column 2 is 0.0314, so increasing a waiter's intrinsic sales ability by 1.4 dollars may increase his/her sales by approximately 4%. In addition, *PartySize* is significant and positive across all models because a larger party size should be positively associated with higher sales per check. Control variables *AbilityStDev* and *LagCoworkerAbility*, however, are insignificant, which fails to support H3. The other two control variables of check-level workload, *AvgTables* and $AvgTables^2$, are significantly positive and negative, respectively, which is consistent with Tan and Netessine (2014), who find an inverted U-shaped relationship between workload and waiters' sales performance.

The 2SLS estimation results rely on the validity and the asymptotic consistency of instrumental variable estimators. Hence, we now check both the relevance condition and the exclusion restriction condition. Columns 5 and 6 of Table 3 show the first-stage regression results. Our combined instrumental variables are not "weak", and they should satisfy the relevance condition. When both endogenous variables are regressed, the coefficients of *System* are both positive (0.0735 and 0.1482) and significant, which implies that the implementation of the new scheduling system may have increased the average team ability. The coefficients of the one-week lagged instrumental variables are also statistically significant with expected signs. Finally, the F-statistics for the joint significance of the first-stage estimations are both over 1,000, which is higher than 10, the suggested rule of thumb for weak instruments (Staiger and Stock, 1997).

Just as important as the relevance condition, the exclusion restriction condition should be satisfied for our instrumental variables. First, we conduct Sargan tests of over-identifying restrictions to test the exclusion restriction condition (Kennedy, 2003). The p-values of the Sargan tests are over 0.1 for both models, which suggests that we fail to reject the null hypothesis that the error terms of the structural models are uncorrelated with the instrumental variables. Further, the implementation of the new scheduling system should affect restaurant performance only through labor decisions because it should not affect demand factors. To sum up, our proposed instrumental variables should be valid because they seem to satisfy both the relevance and the exclusion restriction conditions.

5.4 Mechanisms of waiters' Performance Variation

Having argued that peer effects may exhibit inverted U-shaped relationships with sales, we want now to understand the mechanisms of such performance impacts. Although our observational data do not allow us to examine all possible mechanisms, we examine two that may complement existing peer effect studies.

5.4.1 Waiters' Sales Actions

What actions do waiters take in response to their peers' ability? We analyze the number of items sold, which can help us understand the two different actions that waiters generally take to influence sales: cross-selling and upselling. Cross-selling is selling more items, such as desserts or wines, which will increase sales per check. Upselling means selling more expensive items, such as steaks or seafood instead of chicken, which likewise increases sales per check. To delineate these interwoven factors, we first analyze the effect of the coworker ability on the number of items sold during a check, which is a reasonable proxy for waiters' cross-selling efforts. In particular, we use the same 2SLS strategy and the same set of instruments employed in Subsection 5.2 to estimate the effect of *CoworkerAbility* on a new dependent variable, $\log(Items_i)$, which is the logarithm of the number of items sold during check i. We then control for $\log(Items)$ in the sales model to examine the impact of coworker ability on upselling effort.

Table 4a shows the results of waiters' sales actions analysis. In the $\log(Items)$ model (column 1), the co-

Table 3: Check-Level Peer Effects Results

	(1) Sales Estimated by OLS	(2) Sales Estimated by 2SLS	(3) Sales Estimated by OLS	(4) Sales Estimated by 2SLS	(5) First Stage *CoworkerAbility*	(6) First Stage *CoworkerAbility*^{2}
CoworkerAbility	0.0033	0.0217*				
	(0.0022)	(0.0097)				
CoworkerAbility^2	-0.0023**	-0.0075*				
	(0.0008)	(0.0037)				
RelativeAbility			0.0012	-0.0188		
			(0.0020)	(0.0194)		
RelativeAbility^2			-0.0014***	-0.0274*		
			(0.0002)	(0.0133)		
OwnAbility	0.0301***	0.0314***	0.0331***	0.0447***	-0.0867***	-0.0605***
	(0.0007)	(0.0010)	(0.0023)	(0.0103)	(0.0007)	(0.0021)
PartySize	0.2743***	0.2744***	0.2742***	0.2730***	-0.0022**	0.0025
	(0.0008)	(0.0008)	(0.0008)	(0.0011)	(0.0008)	(0.0025)
AbilityStDev	-0.0035	-0.0048	0.0002	0.0929	0.2518***	0.6465***
	(0.0020)	(0.0035)	(0.0020)	(0.0493)	(0.0019)	(0.0057)
LagCoworkerAbility	-0.0006	-0.0100	-0.0006	-0.0057	0.4720***	0.0719***
	(0.0021)	(0.0054)	(0.0020)	(0.0054)	(0.0019)	(0.0056)
AvgTables	0.0560***	0.0565***	0.0557***	0.0542***	0.0171***	0.0805***
	(0.0045)	(0.0045)	(0.0040)	(0.0043)	(0.0042)	(0.0124)
AvgTables^2	-0.0123***	-0.0124***	-0.0123***	-0.0123***	-0.0019*	-0.0065**
	(0.0008)	(0.0008)	(0.0007)	(0.0008)	(0.0008)	(0.0022)
StoreTables	0.0028***	0.0027***	0.0028***	0.0021***	-0.0027***	-0.0220***
	(0.0002)	(0.0002)	(0.0002)	(0.0004)	(0.0002)	(0.0006)
System					0.0735***	0.1482***
					(0.0058)	(0.0170)
LWTeamAbility					0.2054***	0.0194**
					(0.0020)	(0.0060)
LWTeamAbility^2					-0.0024**	0.2252***
					(0.0009)	(0.0026)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	206,257	201,567	206,257	201,567	201,567	201,567
Prob>Chi-sq	<.001	<.001	<.001	<.001	<.001	<.001
Adjusted R^2	0.437	0.438	0.438	0.398	0.859	0.359

1. Standard errors are shown in parentheses. 2. $*p \leq .05$, $**p \leq .01$, $***p \leq .001$.

efficient of *CoworkerAbility*^{2} is negative and significant (-0.0096), while the coefficient of *CoworkerAbility* is effectively zero, which suggests that coworker ability has an inverted U-shaped relationship with waiters' cross-selling efforts, with an inflection point near the sample mean ($0.02). In other words, as coworker ability increases, waiters first sell more items, but then sell fewer items. In addition, in the log(*Sales*) model conditioned on the number of items sold (column 2), we find that the coefficient of *CoworkerAbility*^{2} is still significantly negative (-0.006), while the coefficient of *CoworkerAbility* is significantly positive (0.0205), hinting that coworker ability has an inverted U-shaped relationship with waiters' upselling behavior, and also suggesting that the inflection point is about ($0.0205/(2\times 0.006) \approx 1.7$) dollars above the sample mean. Putting the two models together, Table 4a(a) suggests that when overall coworker ability is below the sample mean, increasing coworker ability may trigger waiters to expend more effort on upselling and cross-selling. As coworker ability reaches the sample mean, increasing coworker ability may start to stimulate waiters to reduce their cross-selling effort. Although cross-selling effort is reduced, waiters continue to redouble their upselling effort until the coworker ability reaches \$1.7 above the sample mean, at which point upselling effort also starts to fall.

Together with the results shown in Table 3, these results provide insights into the respective sales effects on waiters' cross-selling and upselling activities. From Table 3, we calculate that the optimal coworker ability is about \$1.4 above the sample mean. Given this optimal coworker ability, the average sales may increase by 1.7% because of upselling ($0.0205\times 1.4 - 0.0046\times 1.4^2 \approx 2\%$). However, this optimal coworker ability may reduce waiters' cross-selling effort, which may cause $0.0096\times 1.4 = 1.34\%$ fewer sold items (column 1 of Table 4a). We compute that these 1.34% fewer sold items may cause further sales reduction by $1.34\% \times 0.3093 \approx 0.4\%$, using the coefficient of log(*Items*) in column 2 of Table 4a. In total, Table 4a suggests that the effect of optimal coworker ability should increase sales on average by ($2\% - 0.4\% = 1.6\%$), which is consistent with the estimation in Table 3 (1.6%).

Furthermore, we divide sales ability into cross-selling and upselling abilities to understand how these sub-abilities affect waiters' cross-selling and upselling performance, respectively.[5] Specifically, we replace the sales in Model 1 with the number of sold items to estimate the intrinsic cross-selling ability. In addition, we control for the number of sold items in Model 1 to estimate the up-selling ability. After that, we take the average of these abilities of coworkers to construct the peer effects in terms of cross-selling and upselling abilities (i.e., *CoworkerCSAbility* and *CoworkerUSAbility*, respectively). The results are shown in columns 3

[5] We thank an anonymous reviewer for this suggestion .

and 4. The coefficients of the quadratic terms are both significant and negative (-0.1137 and -0.0042), which suggest that peer effects measured in terms of either cross-selling or upselling ability affect corresponding sales performance in an inverted-U fashion, further supporting our hypothesis.

5.4.2 Table Proximity

As shown in Figure 2, we have information on the assigned tables of each waiter in one of the three restaurants in our study and the locations of these tables. We define an observable waiter if he/she is assigned to a table that is adjacent to any of the tables where the focal waiter works. Then we reconstruct the peer effect variables *AdjacentCoworkerAbility* and *SeparateCoworkerAbility* using the intrinsic abilities (*OwnAbility*) of only those adjacent waiters and separate waiters, respectively, $AdjacentCoworkerAbility_i = \bar{\theta}_{-j}^{adj} = 1/n_1 \sum_{k \neq j} \theta_k^{adj}$, where waiter j handles check i and waiter k (one of n_1 waiters) is observable to waiter j; and $SeparateCoworkerAbility_i = \bar{\theta}_{-j}^{sep} = 1/n_2 \sum_{l \neq j} \theta_l^{sep}$, where waiter l is unobservable (one of n_2 waiters) to waiter j. Mean-centering and replacing these new peer effect variables and their quadratic terms for *CoworkerAbility* and $CoworkerAbility^2$ in Model 2, respectively, we use OLS to test if the peer effect changes because of the waiters' observability. We also include all the adjacent and separate coworkers in one regression.

Table 4b shows the results of the peer effects of observable waiters and unobservable ones. In column 5, where coworkers are observable, the coefficients of the quadratic terms of coworkers' ability are significant and negative (-0.0086), supporting the inverted U-shaped peer effects. However, the coefficients of the coworkers' ability are all insignificant in column 6, which seems to suggest that the peer effect is effectively zero if these coworkers are unobservable, which supports H2. Simply put, it is a case of out of sight, out of mind. These results are qualitatively robust in column 7, where the coefficient of the quadratic term of the observable peers' effects is statistically significant, while the coefficients of non-observable peers' effects are statistically insignificant. Previous work by Mas and Moretti (2009) finds similar evidence that only those cashiers who can directly observe fast coworkers in front of them may make an improvement in their productivity. Although Table 4b provides some evidence that the peer effects are only present if the coworkers are observable at work, it does not preclude that waiters working at table sections far apart may affect each other in common areas such as the kitchen, the drink stations and the registers. Rather, the goal of this analysis is to suggest that waiters working in close proximity may have more significant peer effects, consistent with H2.

Table 4: Mechasnims of Waiters' Performance Variation

(a) Waiters' Sales Actions

	(1) log(*Items*)	(2) log(*Sales*)	(3) log(*Items*)	(4) log(*Sales*)
CoworkerAbility	0.0040	0.0205*		
	(0.0106)	(0.0089)		
CoworkerAbility^2	-0.0096*	-0.0046*		
	(0.0041)	(0.0025)		
CoworkerCSAbility			0.0137	
			(0.0221)	
CoworkerCSAbility^2			-0.1137***	
			(0.0317)	
CoworkerUSAbility				0.0153**
				(0.0056)
CoworkerUSAbility^2				-0.0042**
				(0.0015)
log(*Items*)		0.3093***		
		(0.0019)		
Same controls as in Model 2	Yes	Yes	Yes	Yes
Observations	201,567	201,567	201,567	201,567
Prob>Chi-sq	<.001	<.001	<.001	<.001

1. Controls always included. 2. Standard errors are shown in parentheses. 3. *$p \leq .05$, **$p \leq .01$, ***$p \leq 0.001$.

(b) Table Proximity

	(5) log(*Sales*)	(6) log(*Sales*)	(7) log(*Sales*)
AdjacentCoworkerAbility	-0.0190***		-0.0130
	(0.0057)		(0.0071)
AdjacentCoworkerAbility^2	-0.0086***		-0.0083**
	(0.0020)		(0.0025)
SeparateCoworkerAbility		-0.0025	-0.0025
		(0.0032)	(0.0034)
SeparateCoworkerAbility^2		0.0001	-0.0002
		(0.0010)	(0.0010)
Same controls as in Model 2	Yes	Yes	Yes
Observations	59,399	36,780	35,878
Prob>Chi-sq	<.001	<.001	<.001

1. Controls always included. 2. Standard errors are shown in parentheses. 3. *$p \leq .05$, **$p \leq .01$, ***$p \leq 0.001$.

5.5 Robustness Checks

5.5.1 Hour-Level Analysis

We further conduct an hour-level analysis 1) to provide a robustness check, as previous studies on peer effects analyze this level of aggregation (Mas and Moretti, 2009; Chan et al., 2014a); and 2) to offer practical implications for optimal storewide scheduling as restaurants tend to schedule waiters on an hourly basis.

We define the hour-level dependent variable in terms of hourly average sales per check, $HRAvgSales_{tk} = \frac{TotalSales_{tk}}{HRChecks_{tk}}$ (mean = \$42.49, stdev = \$11.08), where $TotalSales_{tk}$ and $HRChecks_{tk}$ are the total sales of all the checks and the number of checks that open in hour t at restaurant k (mean = 13.07, stdev = 8.82), respectively. We elect to focus on hourly average sales per check instead of hourly total sales for the following reasons. First, the hour-level analysis as a robustness test needs to be comparable with the check-level analysis because our hypotheses are developed based on sales per check. Second, the hourly average sales per check is a sales productivity measure and is therefore immune to demand truncation due to constrained capacity. Admittedly, restaurants care about total sales (sales per check $\times$ volume). Ideally we would need the data for turned-away customers for reasons such as congestion to understand its full implications. Since we do not have such data, we control for *HRChecks* in the model, so the interpretation of the dependent variable is really the average sales per check conditioned on the traffic, which has strong supply-side implications.

The independent variable $TeamAbility_{tk}$ is defined as the average team ability during hour t at restaurant k (equal to $TeamAbility_i$, as defined in Subsection 5.2, and with a mean of \$-0.03 and a stdev of \$1.09). It is then centered around its mean for interpretation purposes. Unlike *CoworkerAbility*, which measures the average ability of the coworkers of a focal waiter, *TeamAbility* assesses the average ability of all the waiters. Accordingly, we exclude *OwnAbility* in the hour-level model. If we still observe an inverted-U relationship between *TeamAbility* and the performance measures, we will find additional evidence for the inverted-U relationship between peer effects and waiters' performance. Moreover, the quadratic specification of *TeamAbility* will imply an optimal average team ability for the storewide sales. For other controls, we change the check-level party size control to $AvgPartySize_{tk} = \frac{\sum_{i \in tk} PartySize_i}{HRChecks_{tk}}$, which is the average party size of the checks during hour t at restaurant k (mean = 2.43 diners, stdev = 0.47 diners). $TeamStDev_i$ is measured every hour, so we simply change its subscript to tk in the hourly model. We also change the one-hour lagged peer effect into *LagTeamAbility,* which is the one-hour lagged variable of *TeamAbility.* We further control for $HRChecks_{tk}$, the store traffic, and then divide it by the number of waiters who processed at least one check in the same hour to create $HRTableLoad_{tk}$ and its quadratic term and adjust for the average individual workload (mean number of waiters per hour = 6.26, stdev = 3.41 waiters; mean of $HRTableLoad_{tk}$ = 1.97 tables per waiter, stdev = 0.69 tables). Finally, we use the same set of time/date/location variables as in

Model 2. We specify our final model as follows:

$$\begin{aligned}
\log(HrAvgSales_{tk}) &= \alpha_0 + \alpha_1 TeamAbility_{tk} + \alpha_2 TeamAbility^2_{tk} + \alpha_3 AvgPartySize_{tk} + \\
&\quad \alpha_4 AbilityStDev_{tk} + \alpha_5 LagTeamAbility_{tk} + \alpha_7 HRChecks_{tk} + \\
&\quad \alpha_8 HRTableLoad_{tk} + \alpha_9 HRTableLoad^2_{tk} + \alpha_{10} Controls_{tk} + \varepsilon_{tk}. \qquad (3)
\end{aligned}$$

We estimate this model by 2SLS using the same instruments as in the check-level analysis. Among our findings shown in Table 5a in the Appendix, the coefficient of *TeamAbility*^{2} is significant and negative (-0.0156), while the coefficient of *TeamAbility* is significant and positive (0.0552). These results suggest that peer effect is likely to have an inverted U-shaped relationship with sales per check, and the optimal team ability to maximize sales is greater than the sample mean, consistent with our check-level results. Interpreting these estimated coefficients, we find that the optimal *TeamAbility* for the entire store is about \$$(0.0552/2 \times 0.0156 \approx 1.76)$ above the sample mean (-\$0.03). In addition, the optimal *TeamAbility* would have increased *HRAvgSales* by $(0.0552 \times 1.76 - 0.0156 \times 1.76^2) \approx 5\%$. This 5% sales lift from optimal team ability composition is the total effect of increasing the average team ability, which includes the direct effect of using higher-ability waiters, and the indirect effect via peer effects. This total effect is quantitatively congruent with our check-level estimation, where we find that optimally increasing every waiter's sales ability by 1.4 dollars will be associated with a 4% direct sales lift and a 1% indirect sales lift via peer effects. Hence, the hour-level analysis results are both qualitatively and quantitatively consistent with our check-level results.

Furthermore, the robust inverted U-shaped peer effects finding also supports the premise of H3. Whereas the coefficient of *AbilityStDev* is insignificant in the check-level sales model (Table 3), the coefficient of *AbilityStDev* is significant and positive in the hour-level analysis (0.0101), which lends further support to H3. The statistical insignificance of *AbilityStDev* at the check level may be because 1) check-level data have many constant values of *AbilityStDev* within the same hour, reducing the variation of the variable and inflating the standard errors, and 2) a team of all average waiters is probably indifferentiable from a mixed team of waiters having the same mean but high and low skills during one shift because of the inverted U-shaped relationship. However, in practice, managers have constraints on the size and the types of workers available to roster. With such constraints, our results suggest that the average heterogeneity of team ability per shift is generally positively associated with performance across shifts. We elaborate on this insight in Subsection 6.1.2.

5.5.2 Moderating Effect of Focal Workers' Ability

Our main results suggest an inverted-U shaped peer effects, which may depend on the focal worker's ability.[6] To address this issue we create two variables to check the interaction effects of the focal worker's ability on the peer effects. First, we use the continuous variable *OwnAbility.* Then, we use a dummy variable *LowAbility,* which is equal to one if the focal waiter's ability is below the sample mean of the ability distribution, and zero otherwise. We use both OLS and 2SLS procedures to estimate the peer effects. Table 5b in the Appendix shows the results of the moderating effect of focal worker's ability. As can be seen, the coefficients of $CoworkerAbility^2$ are all significant and negative across models, consistent with our main results. In addition, all the interaction terms turn out to be statistically insignificant. These results suggest that the inverted-U shaped peer effect is robust, regardless of the focal worker's ability.

5.5.3 Familiarity Weighted Peer Effect

Similar to airline crews and other fluid teams (Huckman et al., 2009), the team composition of every shift at restaurants changes quite substantially over time for reasons such as waiter time preferences and staff turnover. Consequently, during any shift, waiters have familiarity of varying degrees with their coworkers. Familiarity may have implications for worker performance: for example, Huckman et al. (2009) find that team familiarity has a significant positive effect on performance. Just as important, Mas and Moretti (2009) report that the peer effect is insignificant if the coworkers have low schedule overlap with the focal worker. We therefore examine whether the inverted U-shaped peer effect is robust, depending on the degree of

[6]We thank an anonymous referee for this excellent point.

familiarity. Specifically, let's suppose waiters j and k worked together in the same hour when check i started. We define F_{jki} as the number of overlapped hours that waiters j and k had by the hour when check i started, and then we weight *CoworkerAbility* by the familiarity. In other words, $FWCoworkerAbility_i = \bar{\theta}_{-j}^{Familiarity} = 1/n \sum_{k \neq j}(\theta_k \cdot F_{jki})$, where θ_k is the intrinsic sales ability of waiter k (among other n workers) working during the same shift with waiter j. We then replace *CoworkerAbility* and *CoworkerAbility*^{2} in Model 2 with this weighted peer effect measure and its quadratic term and employ 2SLS with the same instruments to test the robustness of the non-linear peer effects weighted by familiarity. We find that the peer effects weighted by familiarity have qualitatively the same inverted U-shaped relationships with sales.[7] Admittedly, F_{jki} is left censored due to data limitation. Consequently the first few months are the most prone to such biases. In order to alleviate this issue, we conduct subsample analysis to focus on the last few months.[8] In particular, we analyze Model 2 using data during the last seven, six, five, four, three and two months, respectively. All of these subsample analyses show consistent inverted U-shaped peer effects, as shown in Table 5c. In addition, we suspect that the potential censoring bias will be quite small because 1) the restaurant industry has one of the highest staff turnover rates; and 2) the effect of familiarity is likely to decrease over time (Huckman et al., 2009).

5.5.4 Subsample Analysis of Using One Restaurant at a Time

The three restaurants in our main analysis may learn how to improve business operations at different rates, thus creating a correlation bias in estimating the peer effects.[9] In order to address this issue we follow the same 2SLS estimation procedure to conduct subsample analysis, using one restaurant at a time. For the two restaurants that did not implement the labor software, we exclude it from the instruments. Table 5d presents the results of the subsample analysis. The coefficients of *CoworkerAbility*^{2} are consistently significant and negative (-0.1614, -0.0357, and -0.023, respectively), supporting H1 that suggests an inverted U-shaped relationship between peer effects and focal workers' sales performance.

[7] The regression results are available from the authors upon request.
[8] We thank the AE for this valuable suggestion.
[9] We thank an anonymous referee for this insightful suggestion.

5.5.5 Alternative Inverted-U Testing: Spline Regressions and Two-Lines Test

Our study is one of the first to examine the inverted U-relationship between peer effects and performance in the workplace. To identify this non-linear relationship in our main analysis, we utilize a commonly used criterion, the significance of the quadratic term of *CoworkerAbility*. Nevertheless, the quadratic specification may suffer from two issues. First, it may create an extreme point even though the true relationship is concave and monotone (Lind and Mehlum, 2010). Second, the quadratic criterion may be limited to the "non-local" assumption, which implies that the fitted dependent variable $\log(\widehat{Sales})$ at a given *CoworkerAbility* $=$ *CoworkerAbility*$_0$ depend heavily on *CoworkerAbility* values far from *CoworkerAbility*$_0$. The first issue should not apply to our analysis because our extreme points are within one and half standard deviations from the sample mean in the sales model. In order to address the second issue, we follow the literature (e.g., Kesavan et al., 2014) to apply spline regressions, choosing one knot that splits *CoworkerAbility* into two equal-sized groups. Then we estimate a spline regression to fit piecewise linear functions of *CoworkerAbility1* (the lower 50%) and *CoworkerAbility2* (the higher 50%). In the same fashion, we also divide the sample into three equal-sized groups to fit three piecewise spline linear functions.

Alternatively, we follow the procedure prescribed in Simonsohn (2016) to conduct a two-line test of the inverted U-shaped relationship. In particular, we first estimate a flexible model of $\log(Sales)$, including only *CoworkerAbility* and *CoworkerAbility*^{2}. Then we identify $\log(\widehat{Sales})_{\text{max}}$, the most extreme internal fitted value, and $\log(\widehat{Sales})_{\text{flat}}$, the set of $\log(\widehat{Sales})$ values within a standard deviation of $\log(\widehat{Sales})_{\text{max}}$. After that, we estimate an interrupted regression setting the breakpoint as the median *CoworkerAbility* value within $\log(\widehat{Sales})_{\text{flat}}$. The *t* statistics associated with the two lines turn out to be 3.88 and -3.73. The final breakpoint should be 3.88/(3.88+3.73)$\approx$ 0.51 percentile of the *CoworkerAbility* within $\log(\widehat{Sales})_{\text{flat}}$, which is essentially the median. Therefore, we use the estimates from the previous interrupted regression as the final estimates of this two-line test.

Table 5e shows the results of the alternative inverted-U testing, including spline regressions and a two-line test. In the two knot model, the coefficient of *CoworkerAbility1* is significant and positive (0.0116), suggesting that as average coworker ability rises, average sales first increase. However, the coefficient of *CoworkerAbility2* is significant and negative (-0.008), implying that as average coworker ability rises further, sales then drop. The results are qualitatively robust in the three knot model in that the coefficients of *CoworkerAbility1* and *CoworkerAbility2* are both significant and positive and the coefficient of *CoworkerAbility3* is significant and negative. Finally, the two-line test results show that the coefficient of the first line is significantly positive and that the coefficient of the second line is significantly negative, corroborating the inverted U-shaped relationship between the peer effects and focal the worker's sales performance.

6 Managerial Implications and Conclusions

6.1 Managerial Implications

The U.S. restaurant industry employs about 13 million (Mill, 2006) workers, the majority of whom are waiters, and suffers from the lowest labor productivity in the service sector (Freeman, 2008). How to manage these waiters with diversified sales skills to optimize financial performance and to improve productivity has become an ever more pressing challenge for restaurant managers facing increasing pressure in a highly competitive industry. Our study provides three main managerial insights into how to manage this ability heterogeneity through optimal scheduling and team composition decisions.

6.1.1 Managing Superstars

On the one hand, high-ability workers are a special asset for restaurant companies. On the other hand, they are prone to high turnover because they can find outside options more easily than low-ability coworkers, hence the business incurs considerable opportunity costs and training costs. How to manage these superstars is a critical question for managers. Using our 2SLS hour-level estimates, we calculate that having an optimal team ability ($1.76 above the sample mean) in our focal restaurants may increase sales by approximately 5%, which includes 3.5% of direct effect from increasing average sales ability and 1.5% of indirect peer effects. This delineation of the total effect of optimal scheduling implies that the value of having higher-ability waiters at work can not only generate more sales from their own tables (direct effect) but also significantly improve the performance of their coworkers (indirect peer effects). Managers may therefore need to reconsider their compensation schemes to retain and reward higher-performing waiters.

Just as important, how to schedule the "superstars" and the "underdogs" in the optimal working group needs careful consideration. For example, in our focal restaurants where the current average team ability is below the optimum to maximize sales conditioned on the volume, hiring better waiters and scheduling them to work more often can certainly boost the current average team ability. However, this can be hard to accomplish in practice. Instead, we propose a heuristic to schedule the "superstars" to work during those shifts that have fewer waiters scheduled. To explain mathematically why this helps, suppose Team 1 has n_1 waiters, while Team 2 has n_2 waiters, and $n_1 > n_2$. Without loss of generality, we assume the average team skill of Team 1 is the same as the average team skill of Team 2. In other words, $\bar{x}_1 = \bar{x}_2 = \bar{x}$. Now suppose in both teams we exhange a low-ability waiter j for a high-ability waiter k, whose skills are x_j and x_k, and $x_j < x_k$. Then the new average skill of Team 1 is $\bar{x}'_1 = \frac{1}{n_1}[(n_1\bar{x} - x_j + x_k]$, while the new average skill of Team 2 is $\bar{x}'_2 = \frac{1}{n_2}[(n_2\bar{x} - x_j + x_k]$. The difference between the two teams, $\bar{x}'_2 - \bar{x}'_1 = (\frac{1}{n_1} - \frac{1}{n_2})(x_j - x_k) > 0$. Hence, scheduling a "superstar" to work in a small team increases average team ability more than scheduling the same worker in a large team. Accordingly, we suggest scheduling the "underdogs" to work during those shifts that have many waiters because the negative contribution of low ability is minimized across other waiters.

Furthermore, our table proximity results (Subsection 5.4.2) imply that peer effects are stronger in close proximity. Hence, when the average team ability happens to be too low, managers may consider placing the few high performers in more visible sections in order to maximize their positive spillovers, which may even exceed the counter-factual 1.5% sales lift due to peer effects.

6.1.2 Polarizing or Mixing: Benefits of Heterogeneity

In the empirical analyses, we find a robust inverted U-shaped relationship between coworkers' sales ability and focal workers' sale performance. Hence, the premise of H3 is confirmed. Consequently, as a rule of thumb for managers to consider in scheduling, we suggest that mixing workers with various skill levels in one shift is likely to be more beneficial for the entire restaurant's sales performance than polarizing the skill levels. To accompany this argument, let's suppose a manager has to roster from a group of workers having ability levels as follows (3, 2, 2, 1, 1, -1, -1, -2, -2, 0, 0, -3) to form four three-person shifts each day. We randomly sample from these workers without replacement to form four groups and repeat this process 500 times to simulate 500 daily rostering decisions. We find that the regression coefficient between the standard deviation of team ability per shift (a heterogeneity measure) and mean performance is approximately 0.9, using an inverted U-shaped performance output function (-0.5×*average team ability*^{2} + 0.5×*average team ability*).

There is a caveat to this suggestion: the effect of team ability heterogeneity is known to depend on the compensation system. According to Chan et al. (2014a), heterogeneity in team ability may improve team performance under team-based incentives, while it may inhibit team performance under individual-based incentives. Although we agree that compensation systems should affect the effect of ability heterogeneity, we find counter-evidence to this effect under individual-based incentives, probably because Chan et al. (2014a) study linear peer effects whereas our setting is a mixture of both individual- and team-based compensation systems as waiters make money from the tables that they are assigned to and need to work as team with coworkers.

6.1.3 Counter-factual Analysis of Scheduling/Rostering Decisions

We develop two approaches to conduct a counterfactual analysis of incorporating the inverted U-shaped peer effects into scheduling and rostering decisions without changing current worker capacity. In the first approach, we adopt a two-step model, solving a scheduling problem first (deciding how many workers needed for each shift) followed by solving a rostering problem (deciding which workers should work in each shift). During the scheduling part, we make the following assumptions: 1) There are two lunch shifts starting at 11am and 12pm, and four dinner shifts starting at 4 pm, 5pm, 6pm and 7pm. The 4pm shift reflects an early dinner shift, while the 7pm shift represents the closing shift. We exclude shifts starting at 1pm, 2pm and 3pm because, in practice, almost no waiters have those shifts, and we want to limit the number of integer decision variables for computational reasons. 2) Each shift k lasts four hours. For example, a person who starts at noon will cover 12pm, 1pm, 2pm, and 3pm. 3) There are 12 hours in a working day from 11am to 10pm. 4) Each worker costs the same. 5) Hourly demand for the number of workers is correlated and simulated from empirical data. Our decisions variables are integers indicating the number of people needed to start one of the six four-hour shifts, i.e., $C_k, k \in \{1, ..., 6\}$. With that, we formulate our scheduling problem as follows:

$$
\begin{aligned}
\text{Max} \quad & \sum_{k=1}^{6} C_k \\
\text{s.t} \quad & \text{Capacity}_t = f_t(C_k) \geq \text{Demand}_t \quad \forall t \in \{1, \cdots, 12\},
\end{aligned}
$$

where the objective function is to minimize the total daily number of people needed to start each shift, subject to the constraint that the worker capacity has to satisfy the demand for labor during each hour t.

After this scheduling step, we solve a rostering problem, where the decision variables as X_{jk} are defined as binary variables indicating whether a waiter j is rostered to work during shift k. The formulation is shown as follows:

$$
\begin{aligned}
\text{Max} \quad & -0.015 AvgHrAbility^2 + 0.055 AvgHrAbility \qquad (4)\\
\text{s.t} \quad & F = \min\Big(\sum_{k=1}^{2} C_k, \sum_{k=3}^{6} C_k\Big)\\
& P_M = \max\Big(0, \sum_{k=1}^{2} C_k - \sum_{k=3}^{6} C_k\Big), \quad P_A = \max\Big(0, -\sum_{k=1}^{2} C_k + \sum_{k=3}^{6} C_k\Big)\\
& \sum_{k=1}^{2} X_{jk} = 1 \quad \forall j \in \{1, \cdots, F\} \cup \{1+F, \cdots, P_M + F\}\\
& \sum_{k=3}^{6} X_{jk} = 1 \quad \forall j \in \{1, \cdots, F\} \cup \{1+F, \cdots, P_A + F\}\\
& \sum_{j=1}^{F+P_M+P_A} X_{jk} = C_k \quad \forall k \in \{1, \cdots, 6\}.
\end{aligned}
$$

In this formulation, F represents the number of full-time waiters, while P_M and P_A indicate the number of part-time waiters who only work in the morning or the afternoon shifts, respectively. The objective function is to maximize the average hourly sales per check, which is a quadratic function of average hourly team ability (*AvgHrAbility*), whose coefficients are estimated from the hourly analysis in the empirical part. To construct the average hourly team ability, we first create a $(F + P_M + P_A)$ by 1 ***Skills*** vector by simulating the skills for each waiter from the empirical skill distribution (a normal distribution with mean = 0.03, standard deviation = 1.5). We then compute the inner products with the decisions variables $\boldsymbol{X}_k$, and take the average across the 12 working hours, i.e., $AvgHrAbility = \frac{1}{12}\sum_k^{12} \boldsymbol{Skills} \cdot \boldsymbol{X}_k$. For the constraints, the full-time waiters have to work exactly one morning shift and one afternoon shift, which implicitly imposes a constraint that no waiter can start more than one shift during any consecutive four hours. The morning part-timers have to do exactly one morning shift, while the afternoon part-timers have to work only one afternoon shift. Finally, the number of waiters rostered during each shift should be the same as the solution to the previous scheduling problem.

Since the number of decision variables is about 100, we cannot enumerate all the possible solutions to find the global optimum of this non-linear integer program. Instead, we employ a heuristic method called genetic algorithm (GA) method in Matlab, which produces high-quality solutions for large-scale problems in a variety of operations research and computer science studies (Ben Hadj-Alouane and Bean, 1997; Vose, 1999; Debels and Vanhoucke, 2007). We then simulate hourly worker demand and worker ability 500 times, each of which represents one average day. We find that when comparing with the current sample mean, the estimated sales improvement from incorporating the quadratic peer effects is approximately 1.31%, with a standard deviation of 1.7%, which is both statistically and practically significant. In addition, we find that the average hourly standard deviation of team ability (a measure for ability heterogeneity) has a mean of 1.41 and a standard deviation of 0.36, which is statistically significantly higher than the current sample mean of 1 with a standard deviation of 0.6. This comparison provides support to our finding that a higher heterogeneity is generally associated with higher sales performance.

Besides solving this two-step scheduling/rostering integer program, we consider an alternative approach to the counter-factual analysis using continuous decision variables X^c_{jk} that represent the number of waiters of skill category j to be scheduled for a particular shift k (the fraction part means the proportion of time that a waiter works in this shift in the long run). The advantage of this alternative approach is that we can perform optimization exactly, at the cost of simplifying decision variables to be continuous. In addition, we can combine both the scheduling and rostering problems into one analysis.

To operationalize the model we make the following assumptions: 1) Unlike the previous integer programming model, there are nine shifts during the day starting at each hour from 11am to 7pm in this continuous model, each of which lasts four hours. We allow more shifts in the continuous model because we want to include more flexible schedules and therefore obtain more robust results since continuous optimization is more efficient to solve than integer optimization. 2) Similar to the integer model, there are 12 hours in a working day from 11am to 10pm and each worker costs the same. 3) We estimate the hourly worker demand from the data and infer the continuous number of workers starting the four-hour shift every hour (i.e., C^c_k). For example, if there are 4.5 waiters needed at 11am and 7.5 waiters at noon, we can infer that 4.5 waiters are scheduled to start the 11am shift and three waiters are scheduled to start the noon shift. 4) We create 12 equally-spaced skill levels from the empirical distribution ranging from -3.26 to 4.79 to represent 12 categories of waiters in terms of their skill levels and define it as a 12 by 1 vector $\boldsymbol{Skills}^c$. We choose 12 categories because 12 is the 95 percentile of the number of waiters working during an hour. We used 14 categories alternatively and the results are qualitatively and quantitatively robust.

With that, we formulate our scheduling problem as follows:

$$
\begin{aligned}
&\text{Max} - 0.015 AvgHrAbility^{c2} + 0.043 AvgHrAbility^{c} \\
&\text{s.t.} \sum_{j}^{12} X^c_{jk} = C^c_k \quad \forall k \in \{1, \cdots, 9\} \\
&\sum_{k}^{9} X^c_{jk} \leq \text{Capacity}_j \quad \forall j \in \{1, \cdots, 12\}.
\end{aligned} \tag{5}
$$

In the formulation, the objective function is similar to Model 4, which maximizes the average hourly sales per check. Unlike Model 4, the objective function coefficients are re-estimated from the hourly analysis by rounding original sales ability to the 12 sales categories. In addition, $AvgHrAbility^c$ remains $\frac{1}{12}\sum_k^{12} \boldsymbol{Skills}^c_k \cdot \boldsymbol{X}^c_k$. For the constraints, the number of waiters rostered during each shift k is equal to the number of waiters needed to start the shift. Furthermore, any category j of waiters is constrained by their capacity estimated in the data. In particular, we tabulate the 12 skill categories in the data in order to obtain the percentage of each category of waiters worked (p_j). After that, we multiply the total worker requirements, $\sum_k^9 C^c_k$, with the p_j to obtain the worker capacity of each category (Capacity$_j$).

Using this approach, we find the sales lift from the sample mean is approximately 2.4% on average (standard deviation =0.0051), which is statistically significant. In addition, we find that the average hourly standard deviation of team ability has a mean of 1.13 with a standard deviation of 0.1, which is statistically significantly greater than the current sample mean of 1 with a standard deviation of 0.6. This result further corroborates the finding in the previous two-step model and our general finding that heterogeneity tends to be positively associated with higher sales.

6.2 Conclusion

This study makes contributions to two streams of literature, peer effects and optimal scheduling/rostering decisions. First, our finding about the non-linear peer effects may reconcile the seemingly conflicting linear peer effects found in earlier studies – the direction of peer effects may depend on the general level of coworkers' ability. Second, by identifying the spillover effects of heterogeneous ability among coworkers, we examine a formerly overlooked assumption for analytical scheduling models that study workers' heterogeneity (e.g., Cezik and L'Ecuyer, 2008; Bhulai et al., 2008; Bard and Wan, 2008). As a preliminary step to bridge this gap between the empirical peer effects studies and the optimal labor decision literature in service operations, we build relatively realistic scheduling/rostering models that consider peer effects. Of course, both models are not perfect and do not account for many other practical constraints. The purpose of this exercise is to provide a counter-factual analysis to show the potential of incorporating peer effects in scheduling/rostering decisions. It is our hope that this paper will generate interests in the analytical service operations community to design a more robust scheduling/rostering algorithm.

Our empirical findings also have implications for scheduling decisions. First, the inverted U-shaped peer effects imply that managers, when faced with a fixed pool of workers, should mix waiters with heterogeneous ability levels during the same shift. Second, we provide a heuristic to schedule high-ability waiters to work during smaller shifts to increase average team ability because the optimal team ability is above the sample mean in our focal restaurants. Third, through a counterfactual analysis, we find that considering the inverted U-shaped peer effects to optimize current waiters' schedules without changing their capacity may increase sales by between one and three per cent, which highlights the value of empirical research for labor decisions (e.g., Campbell and Frei, 2011; Freeman et al., 2015).

It should be noted that this study has several important limitations. First, although our data set is more granular than previous studies on peer effects, we do not observe the social relationships among the coworkers, which have been found in prior work to moderate peer effects (Mas and Moretti, 2009; Bandiera et al., 2009; Chan et al., 2014a). We also lack other service-quality data, such as complete tips data and customer-satisfaction survey data (However, as a robustness check, we examine the tips paid through credit cards, the only tips data available to us, and find the the tips/sales ratio is quite stable, probably because of the strong social norm of tipping in the United States). Since our waiters are a subset of casual dining waiters, one may raise concerns about the generalizability of our findings. We argue that the waiter body under study as a whole is drawn from the same pool as other casual dining restaurants throughout the United States, probably reducing this generalizability concern. Nevertheless, waiters may operate under different incentive schemes in other countries/cultures, which will require additional study. Finally, while we are primarily interested in examining the impact of peer effects on contemporaneous performance in sales, other research may be fruitful in examining how peer effects affect other dependent variables, such as job satisfaction, retention and long-term performance.

References

Abowd, J. M., F. Kramarz, D. N. Margolis. 1999. High wage workers and high wage firms. *Econometrica* **67**(2) 251–333.

Akşin, Z., M. Armony, V. Mehrotra. 2007. The modern call center: A multi-disciplinary perspective on operations management research. *Production and Operations Management* **16**(6) 665–688.

Akşin, Z., S. Deo, J. Jonasson, K. Ramdas. 2015. Learning from many: Partner diversity and team familiarty in fluid teams. London Business School Working Paper.

Albright, S., W. L. Winston. 2014. *Business Analytics: Data Analysis & Decision Making*. Nelson Education.

Angrist, J., A. B. Krueger. 1994. Why do World War II veterans earn more than nonveterans? *Journal of Labor Economics* **12**(1) 74–97.

Argote, L., D. Epple. 1990. Learning curves in manufacturing. *Science* **247**(4945) 920–924.

Argote, L., P. Ingram. 2000. Knowledge transfer: A basis for competitive advantage in firms. *Organizational behavior and human decision processes* **82**(1) 150–169.

Arlotto, A., S. E. Chick, N. Gans. 2013. Optimal hiring and retention policies for heterogeneous workers who learn. *Management Science* **60**(1) 110–129.

Ata, B., J. A. Van Mieghem. 2009. The value of partial resource pooling: Should a service network be integrated or product-focused? *Management Science* **55**(1) 115–131.

Bandiera, O., I. Barankay, I. Rasul. 2009. Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica* **77**(4) 1047–1094.

Bandiera, O., I. Barankay, I. Rasul. 2010. Social incentives in the workplace. *Review of Economic Studies* **77**(2) 417–458.

Barankay, I. 2012. Rank incentives: Evidence from a randomized workplace experiment. Wharton Working Paper.

Bard, J. F., L. Wan. 2008. Workforce design with movement restrictions between workstation groups. *Manufacturing & Service Operations Management* **10**(1) 24–42.

Ben Hadj-Alouane, A., J. C. Bean. 1997. A genetic algorithm for the multiple-choice integer program. *Operations research* **45**(1) 92–101.

Bhulai, S., G. Koole, A. Pot. 2008. Simple methods for shift scheduling in multiskill call centers. *Manufacturing & Service Operations Management* **10**(3) 411–420.

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

Brown, D. J., D. L. Ferris, D. Heller, L. M. Keeping. 2007. Antecedents and consequences of the frequency of upward and downward social comparisons at work. *Organizational Behavior and Human Decision Processes* **102**(1) 59–75.

Buunk, Bram P, Rebecca L Collins, Shelley E Taylor, Nico W VanYperen, Gayle A Dakof. 1990. The affective consequences of social comparison: either direction has its ups and downs. *Journal of personality and social psychology* **59**(6) 1238.

Campbell, D., F. Frei. 2011. Market heterogeneity and local capacity decisions in services. *Manufacturing & Service Operations Management* **13**(1) 2–19.

Carrell, S. E., R. L. Fullerton, J. E. West. 2009. Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics* **27**(3) 439–464.

Carrell, S. E., B. I. Sacerdote, J. E. West. 2013. From natural variation to optimal policy? The importance of endogenous peer group formation. *Econometrica* **81**(3) 855–882.

Cezik, M. T., P. L'Ecuyer. 2008. Staffing multiskill call centers via linear programming and simulation. *Management Science* **54**(2) 310–323.

Chan, D.C. 2016. Teamwork and moral hazard: Evidence from the emergency department. *Journal of Political Economy* **124**(3).

Chan, T. Y., J. Li, L. Pierce. 2014a. Compensation and peer effects in competing sales teams. *Management Science* **60**(8) 1965–1984.

Chan, T. Y., J. Li, L. Pierce. 2014b. Learning from peers: Knowledge transfer and sales force productivity growth. *Marketing Science* **33**(4) 463–484.

Chen, K. P. 2003. Sabotage in promotion tournaments. *Journal of Law, Economics, and Organization* **19**(1) 119–140.

Cohen, S. G., D. E. Bailey. 1997. What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management* **23**(3) 239–290.

Debels, D., M. Vanhoucke. 2007. A decomposition-based genetic algorithm for the resource-constrained project-scheduling problem. *Operations Research* **55**(3) 457–469.

DeHoratius, N., A. Raman. 2007. Store manager incentive design and retail performance: An exploratory investigation. *Manufacturing Service Operations Management* **9**(4) 518–534.

Falk, A., A. Ichino. 2006. Clean evidence on peer effects. *Journal of Labor Economics* **24**(1) 39–57.

Festinger, L. 1954. A theory of social comparison processes. *Human relations* **7**(2) 117–140.

Freeman, M., N. Savva, S. Scholtes. 2015. Gatekeepers at work: An empirical analysis of a maternity unit. London Business School Working Paper.

Freeman, R. 2008. Labour productivity indicators: Comparison of two oecd databases, productivity differentials and the Balassa-Samuelson effect. *OECD Statistics Directorate* .

Gans, N., G. Koole, A. Mandelbaum. 2003. Telephone call centers: Tutorial, review, and research prospects. *Manufacturing and Service Operations Management* **5**(2) 79–141.

Gray, J. V., E. Siemsen, G. Vasudeva. 2015. Colocation still matters: Conformance quality and the interdependence of R&D and manufacturing in the pharmaceutical industry. *Management Science* **61**(11) 2760–2781.

Guimera, R., B. Uzzi, J. Spiro, L. A. N. Amaral. 2005. Team assembly mechanisms determine collaboration network structure and team performance. *Science* **308**(5722) 697–702.

Guryan, J., K. Kroft, M. J. Notowidigdo. 2009. Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics* **1**(4) 34–68.

Hamilton, B. H., J. A. Nickerson, H. Owan. 2003. Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy* **111**(3) 465–497.

Hoegl, M., L. Proserpio. 2004. Team member proximity and teamwork in innovative projects. *Research Policy* **33**(8) 1153–1165.

Hölmstrom, B. 1979. Moral hazard and observability. *The Bell Journal of Economics* 74–91.

Hoorens, V. 1993. Self-enhancement and superiority biases in social comparison. *European Review of Social Psychology* **4**(1) 113–139.

Hopp, W.J., E. Tekin, M. P. Van Oyen. 2004. Benefits of skill chaining in serial production lines with cross-trained workers. *Management Science* **50**(1) 83–98.

Huckman, R. S., B. R. Staats, D. M. Upton. 2009. Team familiarity, role experience, and performance: Evidence from Indian software services. *Management Science* **55**(1) 85–100.

Iravani, S. M. R., B. Kolfal, M. P. Van Oyen. 2007. Call-center labor cross-training: It's a small world after all. *Management Science* **53**(7) 1102–1112.

Itoh, H. 1991. Incentives to help in multi-agent situations. *Econometrica* **59**(3) 611–636.

Itoh, H. 1993. Coalitions, incentives, and risk sharing. *Journal of Economic Theory* **60**(2) 410–427.

Kandel, E., E. P. Lazear. 1992. Peer pressure and partnerships. *Journal of Political Economy* **100**(4) 801–817.

Katzenbach, J. R., D. K. Smith. 1993. *The discipline of teams*. Watertown, MA: Harvard Business Press.

Kennedy, P. 2003. *A guide to econometrics*. Cambridge, MA: The MIT Press.

Kesavan, S., B. R. Staats, W. Gilland. 2014. Volume flexibility in services: The costs and benefits of flexible labor resources. *Management Science* **60**(8) 1884–1906.

Kuziemko, I., R. W. Buell, T. Reich, M. I. Norton. 2014. Last-place aversion: Evidence and redistributive implications. *The Quarterly Journal of Economics* **129**(1) 105–149.

Lapré, M. A., A. S. Mukherjee, L. N. Van Wassenhove. 2000. Behind the learning curve: Linking learning activities to waste reduction. *Management Science* **46**(5) 597–611.

Lazear, E. P. 1989. Pay equality and industrial politics. *Journal of Political Economy* **97**(3) 561–580.

Lind, J. T., H. Mehlum. 2010. With or without U? The appropriate test for a U-shaped relationship. *Oxford Bulletin of Economics and Statistics* **72**(1) 109–118.

Luttmer, E. F. 2004. Neighbors as negatives: Relative earnings and well-being. Tech. rep., National Bureau of Economic Research.

Mandelbaum, A., P. Momcilovic, Y. Tseytlin. 2012. On fair routing from emergency departments to hospital wards: QED queues with heterogeneous servers. *Management Science* **58**(7) 1273–1291.

Mas, A., E. Moretti. 2009. Peers at work. *American Economic Review* **99**(1) 112–145.

Mehrotra, V., K. Ross, G. Ryder, Y. P. Zhou. 2012. Routing to manage resolution and waiting time in call centers with heterogeneous servers. *Manufacturing & Service Operations Management* **14**(1) 66–81.

Mill, R. C. 2006. *Restaurant management: Customers, operations, and employees*. Upper Saddle River, NJ: Prentice Hall.

Moretti, E. 2004a. Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics* **121**(1) 175–212.

Moretti, E. 2004b. Workers' education, spillovers, and productivity: Evidence from plant-level production functions. *American Economic Review* **94**(3) 656–690.

Narayanan, S., S. Balasubramanian, J. M. Swaminathan. 2009. A matter of balance: Specialization, task variety, and individual learning in a software maintenance environment. *Management Science* **55**(11) 1861–1876.

Nembhard, I. M., A. L. Tucker. 2011. Deliberate learning to improve performance in dynamic service settings: Evidence from hospital intensive care units. *Organization Science* **22**(4) 907–922.

O'Connor, E. J., L. H. Peters, A. Pooyan, J. Weekley, B. Frank, B. Erenkrantz. 1984. Situational constraint effects on performance, affective reactions, and turnover: A field replication and extension. *Journal of Applied Psychology* **69**(4) 663–672.

Pisano, G. P., R. M. Bohmer, A. C. Edmondson. 2001. Organizational differences in rates of learning: Evidence from the adoption of minimally invasive cardiac surgery. *Management Science* **47**(6) 752–768.

Roels, G., X. Su. 2013. Optimal design of social comparison effects: Setting reference groups and reference points. *Management Science* **60**(3) 606–627.

Schultz, K. L., T. Schoenherr, D. Nembhard. 2010. An example and a proposal concerning the correlation of worker processing times in parallel tasks. *Management Science* **56**(1) 176–191.

Shafer, S. M., D. A. Nembhard, M. V. Uzumeri. 2001. The effects of worker learning, forgetting, and heterogeneity on assembly line productivity. *Management Science* 1639–1653.

Siebert, W. S., N. Zubanov. 2010. Management economics in a large retail company. *Management Science* **56**(8) 1398–1414.

Siemsen, E., S. Balasubramanian, A. V. Roth. 2007. Incentives that induce task-related effort, helping, and knowledge sharing in workgroups. *Management Science* **53**(10) 1533–1550.

Simonsohn, U. 2016. Two-lines: The first valid test of u-shaped relationships. The Wharton School Working Paper.

Song, H., A. Tucker, K. Murrel, D. Vinson. 2015. Learning from the best: The effects of public relative performance feedback on variability and productivity. Harvard Business School Working Paper.

Staats, B., D. S. KC, F. Gino. 2015. Blinded by experience: Prior experience, negative news and belief updating. University of North Carolina Working Paper.

Staats, B. R., F. Gino. 2012. Specialization and variety in repetitive tasks: Evidence from a Japanese bank. *Management Science* **58**(6) 1141–1159.

Staiger, D., J. H. Stock. 1997. Instrumental variables regression with weak instruments. *Econometrica* **65**(3) 557–586.

Tan, T. F., S. Netessine. 2014. When does the devil make work? an empirical study of the impact of workload on worker productivity. *Management Science* .

Uzzi, B., J. Spiro. 2005. Collaboration and creativity: The small world problem. *American Journal of Sociology* **111**(2) 447–504.

Van den Bergh, J., J. Beliën, P. De Bruecker, E. Demeulemeester, L. De Boeck. 2013. Personnel scheduling: A literature review. *European Journal of Operational Research* **226**(3) 367–385.

Velleman, P. F., D. C. Hoaglin. 1981. *Applications, basics, and computing of exploratory data analysis*. Duxbury Press.

Villas-Boas, J. M., R. S. Winer. 1999. Endogeneity in brand choice models. *Management Science* **45**(10) 1324–1338.

Vose, M. D. 1999. *The simple genetic algorithm: foundations and theory*, vol. 12. MIT press.

Walker, J. R. 2007. *The Restaurant, Study Guide: From Concept to Operation*. New York City, NY: John Wiley & Sons.

Wallace, R. B., W. Whitt. 2005. A staffing algorithm for call centers with skill-based routing. *Manufacturing and Service Operations Management* **7**(4) 276–294.

Ward, A. R., M. Armony. 2013. Blind fair routing in large-scale service systems with heterogeneous customers and servers. *Operations Research* **61**(1) 228–243.

Wiersma, E. 2007. Conditions that shape the learning curve: Factors that increase the ability and opportunity to learn. *Management Science* **53**(12) 1903–1915.

Wooldridge, J.M. 2002. *Econometric analysis of cross section and panel data*. The MIT press.

Appendix

Table 5: Appendix: Various Robustness Checks

(a) Hour-level Team Ability Analysis

	log(*HRAvgSales*)
TeamAbility	0.0552***
	(0.0147)
TeamAbility^2	-0.0156***
	(0.0048)
AvgPartySize	0.3012***
	(0.0036)
AbilityStDev	0.0101**
	(0.0038)
LagTeamAbility	-0.0084
	(0.0073)
HRChecks	0.0032***
	(0.0003)
HRTableLoad	0.0.0678***
	(0.0100)
HRTableLoad^2	-0.0142***
	(0.0020)
Controls	Yes
Observations	14,880
Prob>Chi-sq	<.001

1. Standard errors are shown in parentheses. 2. *$p\leq .05$, **$p\leq .01$, ***$p\leq .001$.

(b) Moderating Effect of Focal Worker's Ability

	Continuous Variable OLS	Continuous Variable 2SLS	Dummy Variable OLS	Dummy Variable 2SLS
CoworkerAbility	0.0041	0.0247**	0.0252	0.0420***
	(0.0023)	(0.0095)	(0.0464)	(0.0103)
CoworkerAbility^2	-0.0032***	-0.0094**	-0.0357*	-0.0084*
	(0.0007)	(0.0035)	(0.0176)	(0.0041)
OwnAbility	0.0306***	0.0340***		
	(0.0009)	(0.0023)		
OwnAbility×*CoworkerAbility*	-0.0008	-0.0026		
	(0.0007)	(0.0018)		
OwnAbility×*CoworkerAbility*^{2}	-0.0007	-0.0029		
	(0.0005)	(0.0023)		
LowAbility=1			-0.0533***	-0.0559***
			(0.0021)	(0.0047)
LowAbility×*CoworkerAbility*			0.0024	-0.0017
			(0.0017)	(0.0034)
LowAbility×*CoworkerAbility*^{2}			0.0009	-0.0029
			(0.0012)	(0.0041)
Same controls as in Model 2	Yes	Yes	Yes	Yes
Observations	206,257	201,567	206,257	201,567
Prob>Chi-sq	<.001	<.001	<.001	<.001

1. Standard errors are shown in parentheses. 2. *$p\leq .05$, **$p\leq .01$, ***$p\leq .001$.

(c) Subsample Analysis of Last *N* Months

	N=7	*N*=6	*N*=5	*N*=4	*N*=3	*N*=2
CoworkerAbility_w	-0.0042	-0.0103	-0.0075	-0.0127	-0.0208	-0.0257*
	(0.0066)	(0.0070)	(0.0077)	(0.0087)	(0.0112)	(0.0127)
CoworkerAbility_w^2	-0.0192***	-0.0199***	-0.0182***	-0.0188***	-0.0196***	-0.0237***
	(0.0028)	(0.0029)	(0.0032)	(0.0036)	(0.0050)	(0.0067)

(d) Subsample Analysis of Using One Restaurant at a Time Results

	Store 1	Store 2	Store 3
CoworkerAbility	-0.4439*	0.0252	0.0654***
	(0.1865)	(0.0464)	(0.0196)
CoworkerAbility^2	-0.1614*	-0.0357*	-0.0231*
	(0.0696)	(0.0176)	(0.0111)
Same controls as in Model 2	Yes	Yes	Yes
Observations	58,983	76,543	66,041
Prob>Chi-sq	<.001	<.001	<.001

1. Standard errors are shown in parentheses. 2. *$p\leq .05$, **$p\leq .01$, ***$p\leq .001$.

(e) Alternative Inverted-U Testing Results

	Two Knots	Three Knots	Two-Lines Test
CoworkerAbility1	0.0116***	0.0077*	0.013***
	(0.0031)	(0.0036)	(0.0035)
CoworkerAbility2	-0.0080**	0.0171***	-0.0126***
	(0.0028)	(0.0048)	(0.0033)
CoworkerAbility3		-0.0148***	
		(0.0034)	
Same controls as in Model 2	Yes	Yes	No
Observations	201,567	201,567	201,567
Prob>Chi-sq	<.001	<.001	<.001

1. Standard errors are shown in parentheses. 2. *$p\leq .05$, **$p\leq .01$, ***$p\leq .001$.