STABLE SCHEDULING INCREASES PRODUCTIVITY AND SALES

The Stable Scheduling Study

STABLE SCHEDULING INCREASES PRODUCTIVITY AND SALES

The Stable Scheduling Study

To Eric Severson,
Roselle Dichoso, and
David Torres,
who took a chance on us –
and Jerome Blazek,
who made it happen.

Distinguished Professor Joan C. Williams,
Associate Professor Susan J. Lambert,
Associate Professor Saravanan Kesavan,
Peter J. Fugiel, Lori Ann Ospina,
Erin Devorah Rapoport, Meghan Jarpe,
Dylan Bellisle, Pradeep Pendem,
Lisa McCorkell & Sarah Adler-Milstein

WORKLIFE LAW
UNIVERSITY OF CALIFORNIA
HASTINGS COLLEGE OF THE LAW

THE UNIVERSITY OF CHICAGO
SCHOOL OF SOCIAL SERVICE ADMINISTRATION

UNC KENAN-FLAGLER BUSINESS SCHOOL

Contents

Acknowledgements

The Stable Scheduling Study was supported by generous grants from the W.K. Kellogg Foundation, the Washington Center for Equitable Growth, the Robert Wood Johnson Foundation, the Institute of International Education in collaboration with the Ford Foundation, the Center for Popular Democracy, the Suzanne M. Nora Johnson and David G. Johnson Foundation, and the Gap.

The study was also made possible thanks to Gap leadership, and to each and every Gap store General Manager, Assistant General Manager, and store leader who participated in biweekly interviews and monthly surveys. We thank you for sharing your insights and experiences, and for trusting us with your critical feedback. The study also would not have been possible without all of the stores' associates, especially those who participated in the monthly associate focus groups and completed the employee surveys.

Thank you also to the team of research assistants who played critical roles in conducting biweekly manager interviews, fielding the employee surveys, and for being the face and the voice of the study in the stores: Nicole Bautista, Jessica Cardott, Stacy Koumbis, Tara Maguire, Hagar Maimon, Michael Park, Ryan Rogulich, Jesse Strecker, Darryl Gras-Partyka, and Whitney Hampton.

Several different researchers took time out of their very busy schedules to provide us comments on this report on a very short timeline: Professor Erin Kelly at MIT Sloan School of Management, Professor Elaine McCrate at the University of Vermont, Heather Boushey at the Washington Center for Equitable Growth, Professor Zeynep Ton at MIT Sloan School of Management, and Professor Serguei Netessine at the Wharton School of Business, University of Pennsylvania.

And as for the Gap, we are grateful to them for a level of commitment that none of the three Co-PIs has ever experienced in our cumulative eighty-odd years of academic research. Gap made significant investments in the study, beginning with a $50,000 seed grant to help fund it. Gap's other outlays included paying the associates for time spent in our associate focus groups and funding the additional hours involved in Targeted Additional Staffing. Gap also invested a lot of staff time: it appointed a project manager at headquarters and allowed us to meet repeatedly with human resources, operations, finance, and other departments; recruit and run associate focus groups; have check-ins with store managers every two weeks for many months; and survey their associates twice. They shared firm data each month for more than a year, which enabled us to assess the impact of the study and develop measures to assess business outcomes. Gap allowed us to access internal company data to a truly extraordinary degree. We also thank Austin Vedder, co-founder and CEO of Shift Messenger, for providing us with detailed transaction data and for incorporating Gap managers' feedback into the app.

We admire Gap's courage and commitment to allowing us to take a sustained, well-funded, evidence-based look at scheduling. Gap holds itself to high standards and acts on its values during a challenging period for the retail industry.

"No matter how you slice it in the retail business, payroll is one of the most important parts of overhead, and overhead is one of the most crucial things you have to fight to maintain your profit margins."

– Sam Walton,
Sam Walton: Made in America (1993)

Executive Summary

Background

Variable schedules are now the norm for part-time workers in a variety of industries including retail, where schedules typically change every day and every week, with three to seven days' notice of the next week's schedule. In recent years, these scheduling practices have come under increasing scrutiny in state attorney general offices, state and local legislatures, and the media.

In retail, unstable schedules for employees have been considered an inevitable outcome of stores' need for profitability. Operations researchers have found that matching labor to incoming traffic is a key driver of retail store profitability (Perdikaki et al., 2012). At the same time, social scientists have studied the deleterious effects of variable schedules on employee wellbeing (Henly & Lambert, 2014). What has been lacking is evidence that schedules in service-sector jobs can be improved in ways that benefit both employers and employees.

What happens in brick-and-mortar retail matters, both to families and to the economy. Even with increasing e-commerce, brick-and-mortar retail remains a cornerstone of America's labor market, employing 15% of the American workforce (Mandel, 2017). Brick-and-mortar retail will probably continue to shrink, but at a slow-to-moderate pace (Mandel, 2017).

Most retailers operate under the assumption that stabilizing employees' schedules would hurt their financial performance because instability is an inevitable outcome of variable demand patterns in retail stores. We tested the validity of this commonly held belief. The goal of our experiment was to determine if it is possible to improve schedule stability without hurting financial performance.

This report describes the first randomized controlled experiment of a multi-component intervention designed to shift schedules in hourly retail jobs toward greater stability.[1] Although prior studies have documented the negative effects of schedule instability and one experiment has tackled schedule unpredictability (Lambert et al., in review), the study described here employs rigorous experimental methods to evaluate an intervention designed to address multiple sources of schedule instability. Our goal was to change an ecosystem in which many practices intersect to undermine schedule stability (Lambert & Henly, 2014).

The multi-component Intervention uniquely targeted four dimensions of schedule stability:

1. Consistency: increasing the consistency of schedules from week to week
2. Predictability: improving the ability of employees to anticipate when they will work
3. Adequacy: giving more hours to employees who want them
4. Input: enhancing employees' input into when they work and when they don't.

The Stable Scheduling Study reflects a partnership between an interdisciplinary team that includes Principal Investigator (PI) Joan C. Williams of the University of California, Hastings College of the Law; Co-PI Susan Lambert of the University of Chicago, School of Social Service Administration; and Co-PI Saravanan Kesavan of the University of North Carolina, Kenan-Flagler Business School; and the Gap, Inc.

The study began with a pretest that lasted from March 2015 to October 2015 in three stores in the San Francisco Bay Area. The pretest was used to develop the components of the Intervention, with input from the stores' managers and the head of human resources, Eric Severson, who introduced us to Gap management and advocated for the experiment. At the end of the pretest, Gap decided to roll out two of our proposed practices to all of their U.S. stores as of October 1, 2015:

1. **Two-week Advance Notice:** All stores were required to finalize and publish schedules two weeks in advance, and
2. **Elimination of On-calls:** All stores were required to cease the practice of scheduling tentative shifts that

The goal of our experiment was to determine if it is possible to improve schedule stability without hurting financial performance.

may be cancelled only a few hours before they are scheduled to start.

The full pilot, which included 28 stores in the San Francisco and Chicago metropolitan areas, ran from November 2015 to August 2016. For the full pilot, we randomly assigned 19 of these stores to the treatment condition and 9 of them to the control condition. During the pilot, both the treatment and control stores continued to implement Two-week Advance Notice and Elimination of On-calls, along with the rest of the company's stores. In treatment stores, we added five additional practices, which we refer to as the "Intervention":

1. **Tech-enabled Shift Swapping:** Through an app called Shift Messenger, associates could swap shifts without requiring manager involvement, and managers could post additional shifts as the need arose.
2. **Stable Shift Structure:** Managers endeavored to increase the consistency of shift start and end times in their store across days of the week.
3. **Core Scheduling:** Managers aimed to improve the consistency of their associates' shifts (days and times) from week to week.
4. **Part-time Plus:** Managers offered a core team of associates a soft guarantee of 20 or more hours a week.
5. **Targeted Additional Staffing:** The research team analyzed store data to identify which stores would be likely to increase their sales by adding additional staff to the sales floor at consistent specified times, and these stores received additional staffing hours at no cost to the store budget.

Major findings

- **Consistency, predictability, and worker input increased.** The size of the increase in these dimensions of schedule stability seems modest, but we have few benchmarks for comparison. Even these modest increases in scheduling stability delivered pronounced improvements in business outcomes.
- **Adequacy of work hours did not increase for most associates.** Part-time Plus associates saw an increase in hours during the intervention period, but the average associate did not. Compared to their scheduled work hours, associates in treatment stores averaged a bit less work (20 minutes) during intervention weeks and an hour less work during weeks they used Shift Messenger.
- **Experienced associates may have benefited more from the shift to stability than inexperienced ones.** Associates assigned Part-time Plus had longer average seniority than other part-time associates and retention increased among seasoned associates in treatment stores.
- **Part-time Plus delivered more hours to a core group of associates but not at the expense of coworkers.** Part-time Plus consistently delivered a minimum of 20 hours a week to a core group of associates, increasing the adequacy of work hours for this subset of associates. Initial analyses indicate that the additional hours garnered by Part-time Plus associates did not result in fewer hours for other part-time associates.
- **Stable scheduling sharply increased median sales** by 7% in treatment stores during the intervention period – a dramatic increase in an industry in which companies often work hard to achieve increases of 1–2%. We estimate that shifting to more stable schedules, over a 35-week period, yielded $2.9 million in increased revenues for Gap in the 19 treatment stores.
- **Return on investment was high.** The only out-of-pocket cost associated with stable scheduling was roughly $31,200 spent on Targeted Additional Staffing. Even taking into account the additional time spent implementing the full intervention, we expect the ROI in stable scheduling to be high. If stable scheduling were adopted enterprise-wide, however, transition costs might well entail upgrading or replacing existing software systems.

- **Stable scheduling also significantly increased labor productivity** by 5%. Treatment stores generated an additional $6.20 of revenue per hour of labor than did control stores. This increase in productivity was likely driven by improved retention of more seasoned sales associates in stores that shifted to more stable scheduling. (Overall retention remained unchanged.)
- **Fluctuating customer demand is not the primary source of instability.** A widespread assumption – that fluctuations in customer demand make stable schedules impossible in retail – is inaccurate. Only 30% of the variability in weekly payroll hours was explained by changes in traffic from week to week. Store managers identified the following key sources of headquarter-driven instability: inaccuracies in shipment information, last-minute changes in promotions, and visits by corporate leaders.

Stable scheduling sharply increased median sales by 7%.

- **Managers consistently implemented Advance Notice and Elimination of On-calls** in both control and treatment stores. Although at first they expressed apprehension about Two-week Advance Notice and Elimination of On-calls, most managers quickly adapted to both practices. Managers generally felt that eliminating on-calls added value and most were enthusiastic. As for Advance Notice, some struggled to publish *every* schedule two weeks in advance, yet nearly 90% of schedules were published at least two weeks in advance during the first three quarters of 2016. Ultimately, Gap reduced the Advance Notice policy from two weeks to ten days at the study's conclusion.[2]
- **Employees were generally enthusiastic** about the shift to more stable scheduling and found that it allowed them to plan their lives better, although some felt there was a gap between promise and execution.
- **Store managers' responses to the Intervention were generally positive, with a strong recommendation that other business practices be changed at the same time as scheduling practices.** Many managers spent less time preparing schedules, saw improvements in associate morale, performance, and commitment, and believed that stable scheduling was the right thing to do. On the down side, some struggled to implement the practices due to instability driven by headquarters and by associates.

Common scheduling practices, notes Professor Zeynep Ton of the MIT Sloan School of Management, result in:

> Employees rushing, taking shortcuts, unable or unwilling to spend time with customers. We will have stores with misplaced products, messy shelves, long checkout lines, a lot of wasted inventory, and a lot of unhappy customers (Ton, 2014, p. 158).

Retailers, she noted in 2014, did not know for sure how much all this costs them. Our findings will help them better assess these hidden costs.

Most of the algorithms being employed by workforce management vendors have their genesis in operations research targeted towards better scheduling of machines in factories. But workers are not machines. While lean, demand-driven scheduling practices have enabled the manufacturing industry to improve efficiency and profitability, when applied to people the resulting instability can negatively impact key business outcomes.

This study contests the widespread conventional wisdom that unstable schedules are impossible and/or uneconomic in today's fast-paced, low-profit, brick-and-mortar retail environment. A growing movement led by Professor Ton documents one path to more schedule stability and higher profits: the "Good Jobs Strategy" (Ton 2014, 2017). Our experiment suggests that even retailers that are unwilling or unable to adopt the entire Good Jobs Strategy nonetheless can offer employees more schedule stability. We found that giving employees more schedule stability increased both sales and labor productivity, signaling a high return on investment. More stables schedules are a win-win for both retailers and their employees.

This study also provides new insight into the change management strategy companies should use to implement more stable scheduling. We found that a lot of the instability of schedules in retail stems from decisions made at headquarters – decisions about promotions, shipment, and leadership visits. For maximum effect, stable scheduling initiatives should use an integrated strategy that involves not just human resources but also marketing, supply chain, finance and the leadership team.

"It is unclear why retailers who invest millions of dollars to drive traffic into the stores through marketing activities would not invest sufficiently in labor planning to ensure that the incoming traffic is converted to sales."

– Kesavan and Mani (2015)

Introduction

Retail scheduling often imposes burdens on employees

Many workers in today's hourly jobs face schedules that create instability in multiple ways, from inconsistent hours and short notice to schedules decided with little employee input (Boushey & Ansel, 2016; Clawson & Gerstel, 2014; Lambert et al., 2014; McCrate, 2017; Williams & Boushey, 2010). In a 2016 representative survey of US residents, 80% of hourly workers reported fluctuations in the number of hours they work week to week, fluctuations that averaged 38% of their usual weekly hours.[3] Most of the variation in the number of hours worked was driven by employers, not employees; only 17% of hourly workers said that they determine the number of hours they work (either independently or within employer guidelines).

Compounding the problem is that schedules are not just inconsistent but also unpredictable. Fully 40% of these hourly workers reported knowing their schedule a week or less in advance, with over a quarter (28%) reporting three days or less advance notice. This makes planning difficult for anyone, but particularly for adults with caregiving responsibilities, for students who can ill afford to miss classes or study time, and for those juggling multiple jobs.

Unstable scheduling is widespread in food service, health care, and other industries but is particularly common in retail (Lambert et al., 2014; Carré & Tilly, 2017). The vast majority (87%) of early-career retail employees (age 26–32) in a national survey reported fluctuations in weekly work hours that averaged 48% of their usual hours. In addition, 50% reported a week or less advance notice, and 44% said their employer determines the timing of their work unilaterally (Lambert et al., 2014; Golden, 2016).

Exacerbating the problems posed by these sources of instability is the inadequacy of hours. Full-time work is increasingly scarce in retail, and part-time schedules often offer few hours, resulting in small paychecks and economic hardship. A recent survey of 1,100 front-line retail workers by the Center for Popular Democracy found that 47% of workers surveyed were working part-time, and that 61% of full-time workers reported starting as part-timers (Corser, 2017). A growing share of this part-time employment is involuntary, meaning workers that desire full-time work can only find part-time jobs. Retail and hospitality contributed 63% of the growth of all part-time employment in the US since the recession of 2007 and over half (54%) of the growth in involuntary part-time employment (Golden, 2016). Almost half (45%) of part-time sales associates at one retailer wanted to work more hours at the company, 31% worked more than one job, and a quarter reported difficulty living on their household income (Henly & Lambert, 2015). This is not surprising since median annual earnings for full-time, full-year workers in retail are $31,980, far below the median earnings for all full-time, full-year workers in America ($44,720) (Bureau of Labor Statistics, 2017).

Unstable scheduling thus has multiple dimensions that intersect to undermine retail employees' personal responsibilities, economic security, and wellbeing. The Stable Scheduling Study developed and evaluated a multiple-component intervention that uniquely targeted four aspects of work schedules for improvement – consistency, predictability, adequacy, and input.

Retail scheduling imposes burdens on employers, too

While there is a gap in the literature on how unstable scheduling affects employers – a hole this study helps fill – a body of research examines the related practice of "lean scheduling." Lean scheduling, which is scheduling for the bare minimum of labor hours, reflects retailers' desire to control costs, with labor estimated to account for 85% of controllable costs in retail stores (Kesavan et al., 2014). Several studies in the last decade question the assumption that the savings offered by "lean scheduling" outweigh its costs (Fisher et al., 2017; Fisher et al., 2006; Ton, 2009, 2014; Netessine et al., 2010; Mani et al., 2014; Kesavan & Mani, 2015).

Many studies document that retail is plagued by poor execution of business processes that can depress sales (Salmon, 1989; Raman et al., 2001a, 2001b; DeHoratius & Raman, 2003; Ton & Raman, 2010; Corsten & Gruen 2003; van Donselaar et al., 2006; Fisher et al., 2006; Netessine et al., 2010). Lean scheduling affects execution directly through customer service and indirectly through back-end operations. Customer service is affected when employees don't know where something is or when would-be customers can't find someone to help them, or help them fast enough. Backend operations suffer when "process conformance" is affected: staff can't execute prescribed processes such as replenishment, doing markdowns correctly, keeping the store looking tidy and inviting, and putting stock in the right place so the next person can find it (Ton, 2009).

Operations research critiques the "myopic mindset" that causes retailers to spend $17.2 billion on advertisements to drive customers into stores, only to embrace a labor model that undermines stores' ability to convert that traffic into sales (Perdikaki et al., 2012). "It is unclear why retailers who invest millions of dollars to drive traffic into the stores through marketing activities would not invest sufficiently in labor planning to ensure that the incoming traffic is converted to sales," note operations researchers (Kesavan & Mani, 2015, p. 33). This body of research has found that low conversion rates have a ripple effect. A failure to convert traffic into sales, for example when a customer leaves because no one is available to find a garment in her size, depresses store traffic in the future, as the customer decides not to return because of a prior negative experience (Perdikaki et al., 2012).

Lean scheduling reflects "business-school thinking gone wrong," concludes Wharton Professor Marshall Fisher. "In a store," he explains, "what's measurable is the payroll checks a retailer writes every week to its stores' staffs. What's hard to measure is the impact that stores' staffs have on revenue." He concludes:

> This opens the door to self-delusion. Retailers can convince themselves that they can cut payroll by 5% in the last three weeks of a quarter to meet their profit promise to Wall Street and it really won't impact customer service, because there's probably people in the stores not doing anything anyway (Fisher, 2012).

The result of all this is a "vicious cycle" in retail that persists due to cognitive bias, according to MIT Professor Zeynep Ton. "We humans are wired to emphasize the short term at the expense of the long term," she notes. "We know from previous research that when managers are called upon to weigh costs that are obvious and easy to measure against benefits that are indirect and not immediately felt, they tend to pay too much heed to the obvious costs and make suboptimal decisions" (Ton, 2014, p. 159). The money saved by cutting labor costs is immediate, and immediately visible. The potential benefits of adding additional labor or improving schedule stability are both longer term (e.g. improved sales) and less concrete (e.g. improved customer service). So we would expect what researchers have found: that short-term savings are over-valued while longer-term benefits are undervalued. What we found at Gap is reflective of larger trends in retail that have been documented in study after study.

"When managers are called upon to weigh costs that are obvious and easy to measure against benefits that are indirect and not immediately felt, they tend to pay too much heed to the obvious costs and make suboptimal decisions."

– Zeynep Ton (2014)

We developed and evaluated a multiple-component intervention that uniquely targeted four aspects of work schedules for improvement – consistency, predictability, adequacy, and input.

Description of the Experiment and Business Context

To test the commonly held belief that it is impossible to shift retail schedules towards greater stability, we designed an experiment that shifted associates in some Gap stores, but not others, towards greater schedule stability and measured the results.

Development of Intervention components

The Stable Scheduling Study began with interviews of store managers in virtually all the Gap stores in the San Francisco Bay Area, led by PI Williams and Whitney Hampton, and the stores in the Chicago area, led by co-PI Lambert and Erin Devorah Rapoport. These interviews shaped the initial package by identifying practices that were likely to matter to employees and be feasible for managers. We tested these practices in three stores in San Francisco. The final intervention package was designed in collaboration with our initial Gap executive sponsor, Eric Severson, and the managers of the pretest stores.

By the end of the pretest, we had developed an intervention with seven components. After completion of the pretest, Gap rolled out two of the pretested practices nationwide:

1. **Two-week Advance Notice:** All stores were required to finalize and publish schedules two weeks in advance, and
2. **Elimination of On-calls:** All stores were required to cease the practice of scheduling tentative shifts that may be cancelled only a few hours before they are scheduled to start.

We knew these changes were coming and designed the additional components of our Stable Scheduling Study Intervention as tools managers could use to complement them. Managers in treatment stores, who elected to participate in the study, committed to implementing five additional components to the best of their ability consistent with business conditions. This is the package we refer to as the "Intervention":

1. **Tech-enabled Shift Swapping:** A mobile app allowed associates to swap shifts and store managers to post shifts that needed to be filled after the schedule was published. The app allowed associates to unload shifts they did not want to work and allowed them to pick up more hours. For managers, the app helped replace the loss of on-call shifts as they could post last-minute shifts or cancel a shift that an associate posted if the manager decided the shift was no longer needed.
2. **Stable Shift Structure:** Establishing standard start and end times for shifts for the store, with the goal of making shifts more consistent and predictable for employees and the entry and exit of employees in the store more consistent and predictable for managers.
3. **Core Scheduling:** Improving the consistency of the days and times individual associates were scheduled to work from week to week.
4. **Part-time Plus:** Giving a core team of associates a "soft guarantee" of 20 or more hours a week, intended to increase schedule adequacy and consistency while offering managers a core staff they can count on throughout the week.
5. **Targeted Additional Staffing:** Giving stores more payroll hours to provide managers with the room to structure greater consistency and predictability into work schedules while increasing the adequacy of hours for sales associates. Co-PI Kesavan analyzed store data to identify which stores would be likely to increase their sales by adding additional staff to the sales floor at consistent specified times, and these stores were allocated additional staffing hours (funded by Gap) at no cost to the store's labor budget.

Throughout the experiment, Gap's scheduling software continued to generate schedules as it always had. When managers implemented the Intervention, they more often than not overrode part or all of the schedules generated by Gap's scheduling software.

Recruitment and random assignment

Participation in the experiment was entirely voluntary on the part of store managers, and we followed strict Human Subjects protocols approved by the Institutional Review Boards at our universities. Key features of the experiment include:

- A total of 28 stores participated in the experiment: 15 in the San Francisco Bay Area and 13 in the Chicago metropolitan area. Managers of two stores declined participation in the experiment.
- The 25 participating stores that were not included in the pretest were randomly assigned to control and treatment conditions; the 3 pretest stores continued in the treatment condition. Managers completed the informed consent process before random assignment so they did not know whether or not their store would be implementing the Intervention when they made their decision to participate in the experiment.[4]
- We assigned a larger proportion of the 28 eligible stores to the treatment (19 stores) than control (9 stores) condition so that we could draw lessons from managers' experiences piloting the Intervention while still taking advantage of the rigor afforded by random assignment.
- Of the 19 treatment stores, 13 received additional staffing hours through Targeted Additional Staffing based on Co-PI Kesavan's analysis of staffing; analyses did not reveal the need for additional staffing in the other 6 treatment stores.
- Roll-out of the experiment occurred between November 2015 and February 2016, beginning in San Francisco.
    - Research staff met one-on-one with all store managers to explain the goals of the study and to complete the informed consent process. Managers whose stores were assigned to the treatment condition then attended a training on the components of the Intervention.
    - Managers in treatment stores committed to implementing all of the Intervention components, as business conditions allowed. All-store meetings were held at each treatment store to formally launch the Intervention. At each store, the research team convened the store managers and all associates to explain the study to

everyone. Roll-out occurred over a several month period in each city, in concert with managers' preferences and business imperatives.[5]

- In San Francisco, the Formula Retail Employee Rights Ordinances, also known as the Retail Workers' Bill of Rights, became operative in October 2015. These ordinances require employers to give employees two weeks' notice of their schedules, give predictability pay for schedule changes, provide employees with a good-faith estimate of their work schedule, and offer additional work hours to current part-time employees before hiring new employees, among other requirements (City and County of San Francisco, 2016). These ordinances affected 3 of our treatment stores and 2 of our control stores.[6]

Data sources

Gap was exceedingly generous in providing us with data from firm systems on workers, hours, and outcomes. Data that provide the basis for our quantitative findings are:

- **Monthly personnel data.** A monthly roster with demographic information on all employees in the stores in the experiment, useful for tracking month-to-month turnover and hiring and to examine variations by employee characteristics such as gender, race, age, and seniority.
- **Time-clock data.** The times each employee clocked in and out each work day, useful for examining the stability and adequacy of workers' hours, both the number and the timing.[7]
- **Weekly posted schedules.** The schedule published by managers, key to measuring the extent to which managers made the move to more stable scheduling practices and to gauging predictability (i.e., the fit between the original schedule and actual worked hours).
- **Detailed sales and traffic records.** Sales and traffic recorded in 15-minute increments, useful for identifying periods of understaffing and estimating the effects of the Intervention on store-level sales.
- **Shift Messenger transaction data.** Transaction data from Shift Messenger that recorded every shift posted and picked up and by whom. Combined with personnel data, we examined what shifts were posted and which were filled and by whom.
- **Monthly manager surveys.** Surveys asking about scheduling practices and challenges that managers accessed and completed online each month.
- **Employee surveys.** Baseline and post-intervention surveys of store associates working in participating stores. The response rate to both surveys is around 50%. Overall 760 hourly Gap employees chose to participate in at least one of the surveys, conducted on-line and by mail: 290 responded to both surveys, 246 to the baseline survey only, and 224 to the post-intervention survey only (56% of wave 2 respondents also responded to the wave 1 survey).[8] The survey included questions on their own work schedule, organizational commitment, and health and wellbeing. It was adapted from the survey used in the University of Chicago Work Scheduling Study (Lambert & Henly, 2012; Henly & Lambert, 2014).

In addition to the quantitative data, we also collected in-depth qualitative data, which were coded for themes using the Dedoose qualitative analysis program:

- **Biweekly manager interviews.** A team of research assistants did check-in interviews with store managers (or their designees, typically assistant store managers) every two weeks either on the phone or in person. These interviews started in November 2015 and persisted through August 2016. Research assistants took notes during the interviews, attempting to capture as many verbatim quotes as possible.
- **Focus groups with sales associates.** Focus groups, suggested by Sarah Adler-Milstein, were held with sales associates in treatment stores between January and July 2016. The groups met for two hours a month at various locations convenient to associates across Chicago and the Bay Area. Discussions focused on the implementation of the different Intervention components.

All quotes cited in this report come from recurring biweekly manager interviews and monthly associate focus groups.

Baseline practices and business context

At the beginning of the study, Gap followed scheduling practices that were typical in retail. Store operations calculated the labor budget each store received for

the week based on the sales forecast from the finance department, along with other fixed workload estimates. Then Gap's computerized scheduling system, which Gap refers to as Personnel Management ("PM" for short), produced schedules, staffing the number of hours dictated by the labor budget with associates based on their "availability" – the hours they had told Gap they were available for work. The result was a labor schedule that changed from day to day and week to week, as found in other retailers (Ton, 2014). As in another retail chain in which unstable schedules were studied extensively, store managers felt they were scrupulous about honoring associates' stated "availability," which typically meant that associates were not scheduled to work for times when they said they were not available. Though associates generally can control when they do *not* work, they often cannot control when they *do* work (Lambert et al., 2012).

Once the PM system generated schedules, managers could go in and manually change the schedule, which many did because the shifts generated did not account for skill level or other factors managers found relevant. Managers often felt like the system was working against them by making it impossible to copy and paste schedules week to week. In order to reuse a schedule, a manager had to override the computerized schedule, which took time. Noted one manager wistfully, "I wish PM would remember who I'm plugging in where in terms of skill set." Another said, "At my old job, you could just put in reoccurring shifts and the shift would populate every week." This was not as easy at Gap.

> "It's such a panic every week, so stressful, to make sure that we are not over our hours."
>
> – Store Manager

Another aspect of Gap's scheduling practices that was characteristic of retail is that store managers were penalized if they went over the hours allocated to them. Said one manager, "It's such a panic every week, so stressful, to make sure that we are not over our hours. If we are, we get these emails that go out to the whole district, and if you're over they highlight your store in red."

WHO WORKS AT GAP?

- Gap, Inc. is a global apparel retailer that is the parent of six clothing chains. It employs roughly 135,000 people worldwide (Forbes, 2017). Our experiment was conducted in Gap, Inc.'s namesake clothing chain, Gap. During the timeframe of the experiment, from September 2015 through August 2016, there was a total of 2,331 employees in the 28 stores participating in the study. The smallest participating store had a monthly average headcount of just over 10 employees, whereas the largest had a headcount of over 150.
- Of employees in participating stores, 8.3% were in management positions, including assistant managers, specialized managers, and general managers.
- The vast majority of associates (88%) were sales associates, with the remaining 12% made up of associate interns, stock associates, and lead associates. Only 5% of associates worked full-time.
- About one-third of the associates had worked for the company for less than six months, another third had a tenure between six months and two years, with the remaining third working in the stores for over two years.
- Across all the stores, one-third identified as white, one-quarter as African-American, and one-quarter as Hispanic. In addition, 14% identified as Asian and 2% as Native American.
- Three-quarters of associates were female.
- About one-quarter of the associates were under the age of 20, with another half between the ages of 20 and 29, and the remaining quarter over the age of 30.

A prior case study of another retailer found that 83% of store managers surveyed reported that staying within one's hours budget was "very important" – more important than having the right mix of skills in the store (60%) (Lambert et al., 2012). Staying within your hours' budget also was important at Gap.[9]

As noted, Gap made significant nationwide policy changes at the end of the pretest period resulting in the adoption of Two-week Advance Notice and Elimination of On-calls by all stores. Throughout the experiment, then, both of these practices were standard protocol in both control and treatment stores. Our data indicate that the elimination of on-call shifts took effect immediately. Gap changed the PM system so that managers literally could not schedule on-call shifts, swiftly terminating this practice. Posting schedules two weeks in advance took a bit longer to become standard operating procedure, and our data indicate that three stores that experienced chronic manager turnover had very low compliance throughout the experimental period. Many managers had concerns that, given business unpredictability, it would be too difficult to post further in advance. Overall, however, regardless of whether stores were in the control and treatment condition and despite managers' early concerns, they overwhelmingly complied with the mandate to post schedules two weeks in advance. Excluding the three outliers, 90% of schedules were published at least two weeks in advance during the first three quarters of 2016, after which our experiment ended.

Excluding the three outliers, 90% of schedules were published at least two weeks in advance during the first three quarters of 2016.

How Stable Did Schedules Get?

Overall, the Intervention improved schedule stability by delivering greater consistency, predictability, and employee input. While the shift towards stability on these three measures is statistically significant, our findings on adequacy are mixed, as described in further detail below.

The effect size of the overall shift towards stability seems quite modest, although we have few benchmarks by which to make this call, given that this is the first experiment to target multiple dimensions of schedule stability (as opposed to just predictability).

In gauging the effects of the Intervention on schedule stability we primarily draw on two types of data: hours scheduled and hours actually worked. Scheduled hours are important because the Intervention focused on changing managers' scheduling practices. Actual hours worked allow us to gauge how different components of the Intervention worked together, given that factors beyond managers' control could influence the hours associates actually ended up working, notably shift-swaps and call-offs (when an employee calls to say they cannot come to work).

To estimate the average effect of the Intervention overall, we compare intervention to non-intervention weeks, given that treatment stores entered the experiment at different times.[10] We also compare stores assigned to control and treatment conditions in each fiscal quarter, regardless of whether treatment stores had yet implemented the Intervention. A third comparison, "as good as it gets," identifies the quarter in which schedule stability peaked in order to estimate the maximum amount of stability delivered to associates over the course of the experiment.[11] We also compare responses of associates in control and treatment stores on the post-intervention survey.[12]

1. Consistency

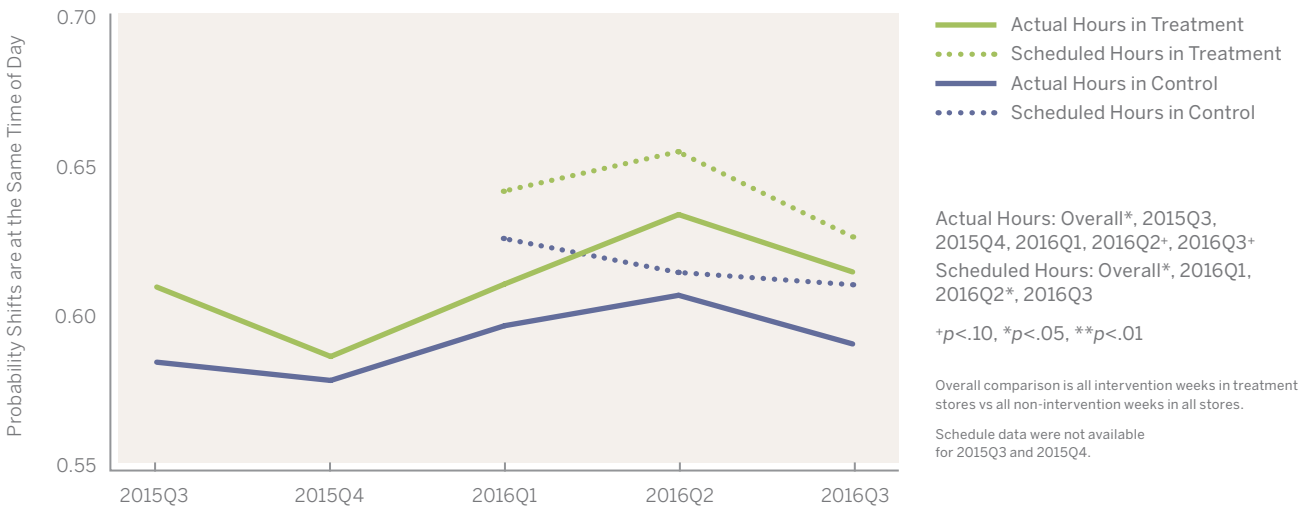
Our first measure of schedule stability captures the consistency of associates' work hours across days and weeks. We examine the consistency of the time of day, shift start and end times, day of the week, and number of weekly work hours. The first three consistency measures are on a 0.0–1.0 probability scale, with higher numbers indicating more consistency.[13] Variation in the number of weekly work hours is measured with a standard deviation. With one exception, which we will explain below, differences between control and treatment stores on schedule consistency are not statistically significant at the start of the intervention period (2015-Quarter 3), adding confidence that the Intervention increased schedule consistency.

Time of day

Figure 1 presents the probability that associates in control versus treatment stores would work during the same

Figure 1

Time of Day Consistency: Scheduled and Actual Hours



Actual Hours: Overall*, 2015Q3, 2015Q4, 2016Q1, 2016Q2+, 2016Q3+
Scheduled Hours: Overall*, 2016Q1, 2016Q2*, 2016Q3

+p<.10, *p<.05, **p<.01

Overall comparison is all intervention weeks in treatment stores vs all non-intervention weeks in all stores.

Schedule data were not available for 2015Q3 and 2015Q4.

general time slots across weeks of the experimental period, averaged for each quarter.[14] Note that data on actual hours are available from 2015-Quarter 3 through the entire experimental period, but we only gained access to published schedules in 2016.[15]

Shift times became more consistent. Overall, associates received schedules with slightly more consistent shift times during weeks the Intervention was in effect ($p<.05$). For example, a part-time associate with an average of 16 shifts per month could expect nearly 10

Below are 4 weeks of published schedules for 2 individual part-time associates who worked at Gap during the intervention period. Total weekly hours are shown on the right-hand side while consistency measures are listed below.

ASSOCIATE A: WEEKLY SCHEDULES

S	M	T	W	T	F	S	Hours
	5 HRS 6:00–11:00						5
	5 HRS 6:00–11:00	5 HRS 6:00–11:00					10
	7 HRS 6:00–13:00			4 HRS 6:00–10:00			11
	7 HRS 6:00–13:00			5 HRS 6:00–11:00			12

Measure	Value
Consistency of scheduled day of week pattern	0.38
Consistency of scheduled start times	1.00
Consistency of scheduled end times	0.43
Consistency of scheduled shift block	1.00

Associate A is consistently scheduled for shifts that start at 6 a.m. and have their midpoint in the early morning (6–10 a.m.), which means the probability of being scheduled for the same time of day and start time is 100%. Associate A's end times and days of the week are less consistent, resulting in a 43% probability of being scheduled for the same end time and a 38% probability of being scheduled for the same days of the week.

ASSOCIATE B: WEEKLY SCHEDULES

S	M	T	W	T	F	S	Hours
4 HRS 10:00–14:00	4 HRS 8:00–12:00				5.5 HRS 9:30–15:00	4 HRS 9:00–13:00	17.5
		7.5 HRS 8:00–15:30		4.5 HRS 8:00–12:30	5.5 HRS 14:00–19:30	4.75 HRS 10:00–14:45	22.25
	4.5 HRS 5:00–9:30	4.5 HRS 15:30–20:00				4 HRS 15:30–19:30	13
	4 HRS 10:00–14:00	4.5 HRS 5:00–9:30			4.75 HRS 8:00–12:45	6 HRS 9:00–15:00	19.25

Measure	Value
Consistency of scheduled day of week pattern	0.25
Consistency of scheduled start times	0.19
Consistency of scheduled end times	0.16
Consistency of scheduled shift block	0.36

Associate B has much less consistent schedules, starting and ending at different times, at different times of day, and different days of the week. The probability that Associate B is scheduled for the same time of day (e.g. afternoon) is 36%, while the probability of being scheduled for the same start or end time is less than 20%. Because there are only 4 weeks of schedules in this example, the probability of being scheduled for the same days of the week is 1 in 4.

shifts per month to be scheduled for the same time of the day during intervention weeks, as compared with about 9 shifts per month for employees during non-intervention weeks. The timing of the shifts associates actually worked tended to be less consistent than their scheduled shifts in both control and treatment stores (see Figure 1), but associates in treatment stores still had more consistent actual hours than associates in control stores. A typical part-time associate working approximately 12 shifts per month could expect nearly 7 shifts to occur during the same time of day each month in treatment stores, as compared to 6 shifts per month in control stores.

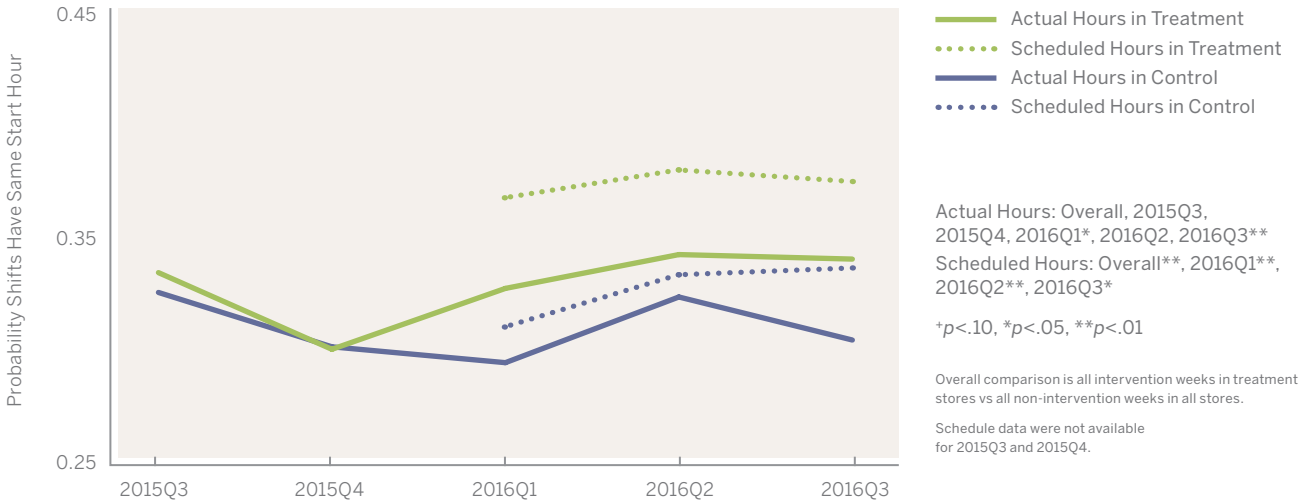
As good as it gets. At its peak (2016-Quarter 2), treatment store associates had a 66% probability of having shifts scheduled at the same time of day, while for control store associates this probability was 62% ($p<.05$) (Figure 1). At that point, nearly 10 in 15 associate shifts had consistent timing in treatment stores, as compared to approximately 9 in 15 shifts in control stores.

Start and end times

Start and end times became more consistent. Scheduled start and end times were more consistent

Figure 2

Start Time Consistency: Scheduled and Actual Hours



Actual Hours: Overall, 2015Q3, 2015Q4, 2016Q1*, 2016Q2, 2016Q3**
Scheduled Hours: Overall**, 2016Q1**, 2016Q2**, 2016Q3*

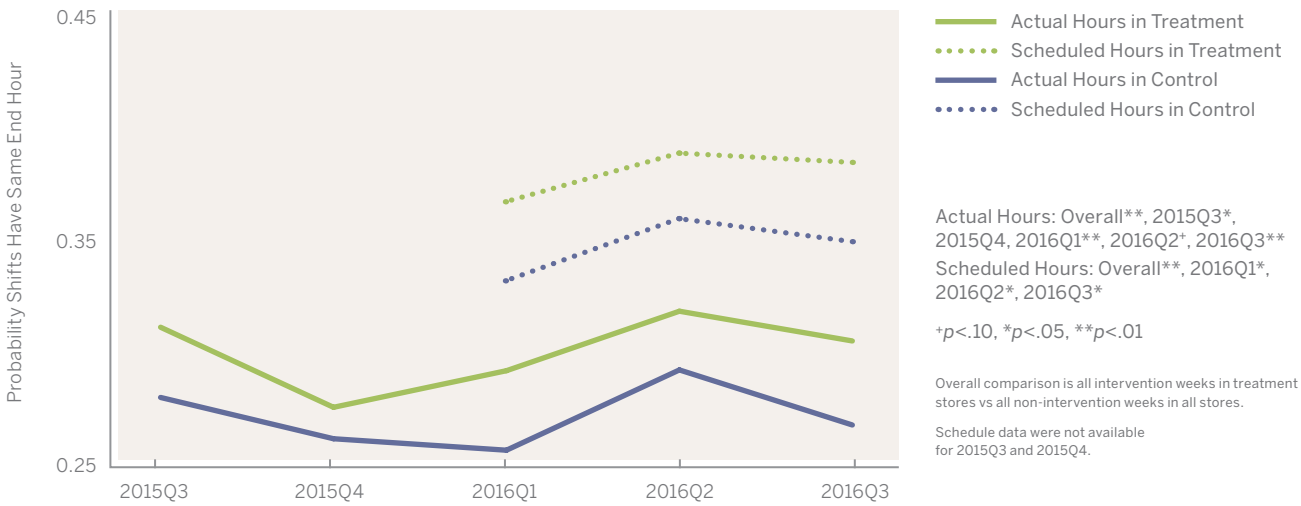
$^{+}p<.10$, $*p<.05$, $**p<.01$

Overall comparison is all intervention weeks in treatment stores vs all non-intervention weeks in all stores.

Schedule data were not available for 2015Q3 and 2015Q4.

Figure 3

End Time Consistency: Scheduled and Actual Hours



Actual Hours: Overall**, 2015Q3*, 2015Q4, 2016Q1**, 2016Q2^{+}, 2016Q3**
Scheduled Hours: Overall**, 2016Q1*, 2016Q2*, 2016Q3*

$^{+}p<.10$, $*p<.05$, $**p<.01$

Overall comparison is all intervention weeks in treatment stores vs all non-intervention weeks in all stores.

Schedule data were not available for 2015Q3 and 2015Q4.

during intervention weeks.[16] The odds of an associate being scheduled for shifts that start or end in the same hour was better than 3 in 9 during intervention weeks but less than 3 in 10 in non-intervention weeks. In terms of actual hours, a typical part-time associate working 12 shifts a month could expect to start work at the same hour about 4 shifts per month whether or not she was in the Intervention. But the same associate would end work at the same time only 3 shifts per month during weeks the Intervention was not implemented, compared with nearly 4 shifts per month during intervention weeks.[17]

As good as it gets. Consistency peaked in 2016-Q2, when the odds of an associate being scheduled for the same start hour is nearly 2 in 5 in treatment stores, as compared with 2 in 6 in control stores (Figure 2). For a typical part-time associate with 3 shifts per week this translates into about one more shift per month (5 shifts with the same start hour in treatment stores as compared with 4 in control stores).

Employees' perceptions of schedule consistency mirror the company data. On the post-intervention survey, respondents from treatment stores were significantly more likely than their counterparts in control stores to agree that they were generally scheduled to begin and end work at the same time each work day.[18]

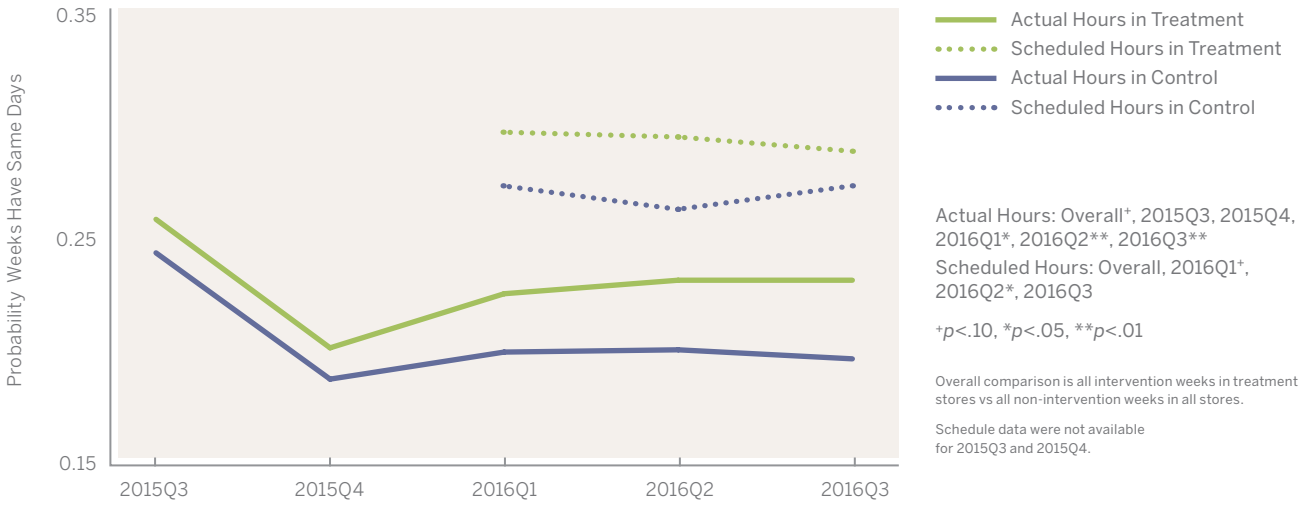
Overall, the Intervention improved schedule stability by delivering greater consistency, predictability, and employee input.

Days of the week

No significant change in consistency of days of week scheduled and little change in days actually worked. The odds of being scheduled for the exact same days of the week were, on average, 3 in 13 during both intervention and non-intervention weeks. On average, an associate worked the exact same days of the week only 2 weeks per quarter.

Figure 4

Days of the Week Consistency: Scheduled and Actual Hours



Actual Hours: Overall+, 2015Q3, 2015Q4, 2016Q1*, 2016Q2**, 2016Q3**
Scheduled Hours: Overall, 2016Q1+, 2016Q2*, 2016Q3

+p<.10, *p<.05, **p<.01

Overall comparison is all intervention weeks in treatment stores vs all non-intervention weeks in all stores.

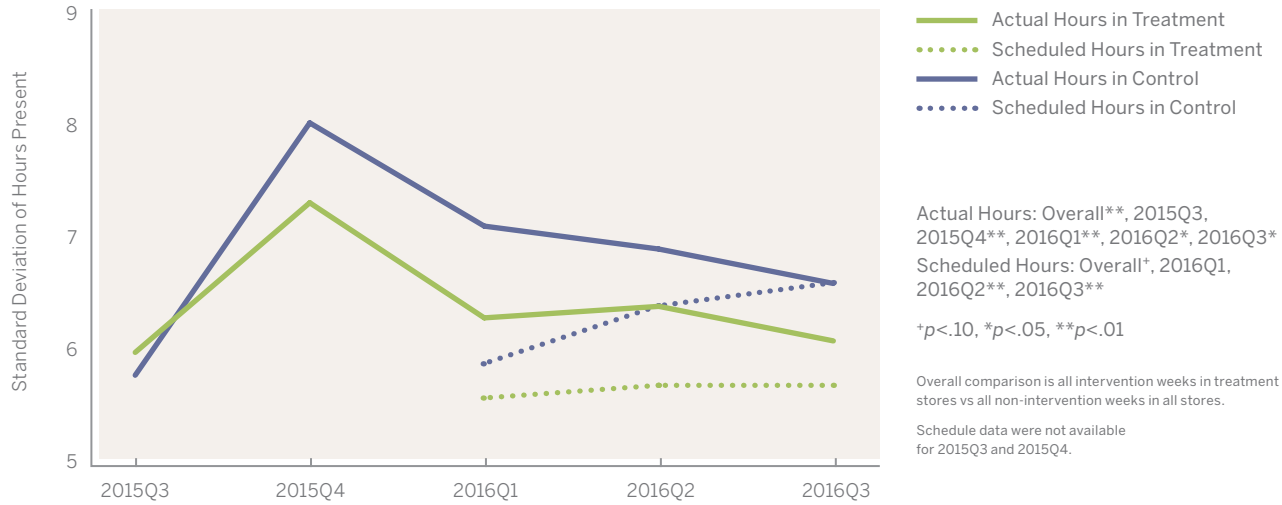
Schedule data were not available for 2015Q3 and 2015Q4.

As good as it gets. In the second quarter of 2016, the consistency of days of the week *scheduled* was greater in treatment than control stores (Figure 4). The typical associate could expect to be scheduled for the same days in about 4 weeks of a (13-week) quarter in treatment stores, as compared with 3 weeks in control stores. The consistency of the days *actually worked* was greater in treatment stores than in control stores during the final quarter of the intervention period, when the odds of an associate actually working the same days of the week were 1 in 4 in treatment stores but only 1 in 5 in control stores (p<.05).

Associates in both treatment and control stores reported more consistency in days scheduled each week than indicated by company data. In the post-intervention survey, 70% of respondents in both treatment and control stores agreed that they are generally scheduled to work the same days each week. There were no significant differences between control and treatment store associates.

Figure 5

Inconsistency: Scheduled and Actual Hours



Actual Hours: Overall**, 2015Q3, 2015Q4**, 2016Q1**, 2016Q2*, 2016Q3*
Scheduled Hours: Overall+, 2016Q1, 2016Q2**, 2016Q3**

+p<.10, *p<.05, **p<.01

Overall comparison is all intervention weeks in treatment stores vs all non-intervention weeks in all stores.

Schedule data were not available for 2015Q3 and 2015Q4.

Variation in number of weekly hours (inconsistency)

Variation in weekly work hours decreased when looking at total number of hours worked. Looking at the number of weekly hours *actually worked*, treatment and control stores begin at roughly equal levels of inconsistency in 2015-Q3, with a standard deviation of 6 hours, but then diverge throughout the intervention period (Figure 5). Looking at *scheduled* weekly hours, the difference between control and treatment stores is not significant in 2016-Q1. But by the end of the experiment in 2016-Q3, the standard deviation in scheduled weekly hours is nearly 7 hours per week in control stores but fewer than 6 hours in treatment stores. For a typical part-time associate who works 3 shifts and 15 hours per week, this difference would add up to an additional gain or loss of a full shift every 6 weeks in control stores as compared with treatment stores. Inconsistency in the total number of weekly hours associates were *scheduled* to work increased in control stores but not in treatment stores across the three quarters for which we have schedule data.

Associates in both treatment and control stores reported more consistency in number of weekly hours than indicated by data from firm systems. In the post-intervention survey, 69% of respondents overall agreed that they are generally scheduled to work the same number of hours each week (differences between respondents in treatment and control stores were not significant).

2. Predictability

Predictability concerns how closely associates' actual hours align with the hours they were scheduled to work. We combine data on the original published schedules with actual work hours to measure predictability in timing (start and end times) and number of weekly work hours. Because our measures of predictability draw on our published schedules data, our analyses are restricted to January through August 2016. These analyses do not distinguish between employee-driven and manager-driven schedule changes.

Predictable timing

Table 1 presents the percentage of shifts during intervention and non-intervention weeks when the timing of a worked shift did not match the timing on the published schedule by a difference of 15 minutes or more. The lower the percentage, the greater the predictability.

Start times were more predictable than end times during both intervention and non-intervention weeks. End times are more unpredictable than start times, because it is more common for associates to leave work early or to stay later than scheduled as compared to being asked to come in earlier or later than originally scheduled.

The timing of shifts, especially end times, was more predictable during intervention than non-intervention weeks. A significantly lower percentage of

Table 1

Predictability in Shift Timing: Percentage of Shifts Changed

Measure	Intervention weeks	Non-intervention weeks	*p*
Percentage of shifts with change at start time	18%	17%	ns
Percentage of shifts with change at end time	34%	40%	<.01
Percentage of shifts with any change*	41%	46%	<.01

*This includes a mismatch between scheduled and worked hours at the start and/or end of a shift.

shifts were changed from the original schedule during intervention (41%) versus non-intervention (46%, $p<.01$) weeks. These differences were driven primarily by fewer changes to the ending time for associates in treatment stores, including significantly fewer shifts extended at the end, cut at the end, or moved back.

As good as it gets. Predictability peaked in 2016-Quarter 3 (not in table), when 38% of shifts in treatment stores changed at either the start or end by at least 15 minutes, compared to 43% of all shifts in control stores ($p<.05$).

The Intervention does not seem to have undermined the predictability generated by Two-week Advance Notice. The additional predictability during intervention versus non-intervention weeks suggests that the other Intervention components, including the shift-swap app, did not undermine the predictability produced by publishing schedules two weeks in advance, which occurred in both control and treatment stores. Associates' survey responses support this conclusion. There were no significant differences between the level of predictability reported by control and treatment store respondents. Two-thirds (66%) of associates responding to the post-intervention survey agreed that they could easily anticipate the time they would begin work each workday, and 62% agreed they could anticipate what time each workday would end.

Predictable number of hours

Associates worked less than originally scheduled during intervention weeks versus non-intervention weeks. During intervention weeks associates worked about 20 minutes *less* than they were originally scheduled, but during non-intervention weeks, associates worked about 5 minutes *more* than they were originally scheduled (intervention -0.27 hours vs. non-intervention 0.13 hours, $p<.05$). The use of the Tech-Enabled Shift Swapping tool may help explain at least part of this difference, as we discuss below.

Part-time associates in treatment stores reported greater predictability in number of weekly hours. Among part-time associates, 71% of respondents in treatment stores agreed that they could easily anticipate how many hours they would work each week, compared to only 63% of part-time associates in control stores ($p<.01$).

3. Adequacy

As intended, Part-time Plus consistently delivered a minimum of 20 hours a week to a core group of associates, increasing the adequacy of work hours for this subset of associates. On average, though, the Intervention did not increase the average number of hours worked by part-time associates, and about half of associates in both treatment and control stores indicated that they would like to work more hours at Gap. Initial analyses indicate that the additional hours garnered by Part-time Plus associates did not result in fewer hours for other part-time associates.

Part-time Plus

Part-time Plus increased the proportion of weeks associates worked 20 or more hours. Associates were scheduled to work 20 or more hours in 70% of the 8 weeks following designation as Part-time Plus, compared to 56% of the preceding 8 weeks ($p<.01$).[19] These associates also picked up additional hours after the schedule was posted, working a minimum of 20 hours in 76% of the weeks subsequent to Part-time Plus status, compared to 56% of the weeks prior to Part-time Plus.

Part-time Plus associates' work hours increased. In the 8 weeks after being designated Part-time Plus, associates were scheduled for an average of 25 hours per week, compared to 21 hours in the prior 8 weeks. Looking at actual hours worked, associates averaged 26

hours in the 8 weeks subsequent to entering Part-time Plus status, compared to an average of 21 hours in the prior 8 weeks ($p<.05$).

Only a small percentage of associates were designated Part-time Plus. In total, 77 employees were designated Part-time Plus by treatment store managers, ranging from 1 to 14 associates in a store.[20] The percentage of part-time associates on Part-time Plus status ranged from 2% to 41% (average 14%; median 4%) of a store's part-time workforce, with no obvious relationship to store size. Associates securing Part-time Plus status were significantly older and had greater seniority than other part-time associates, but they were no more likely to have self-identified as male or as white than other part-time associates.

Adequacy of weekly work hours

The Intervention did not increase number of work hours for the average hourly associate. In fact, associates were scheduled for 54 minutes more in non-intervention weeks than in intervention weeks overall (17.7 vs. 16.6 hours, $p<.05$), and in 2016-Quarter 2 when the Intervention was at its peak, associates worked 90 minutes less per week in treatment stores than in control stores (19.9 hours control vs. 18.3 hours treatment, $p<.05$).[21]

Part-time associates want more hours. About half of part-time associates who participated in the post-intervention survey reported that they would prefer to work more hours at Gap, in both treatment (48%) and control (53%, difference not significant) stores, and a third (33%) of part-time respondents indicated that they would prefer to work full-time if approved to do so by management. Associates with Part-time Plus status were even more likely to prefer additional hours of work: 58% reported preferring more hours at Gap, with 52% indicating they would prefer to move to full-time. This is consistent with the results of a study of another apparel retailer; 45% of part-time associates reported they would prefer to work more hours for more pay (Henly & Lambert, 2015).

4. Input

Shift Messenger, the Tech-Enabled Shift Swapping tool we selected, was adopted with the dual goal of increasing employees' input into schedule changes and providing new opportunities to pick up additional hours of work. Our data suggest that the app increased schedule input but that, on average, the app did not translate into additional hours of work.

Many employees used the shift-swap app to gain control over their work hours. The majority (62%) of part-time hourly associates in the stores that adopted Shift Messenger had posted or picked up a shift during the approximately 30 weeks it was deployed. Although use was especially high among associates under 21 (70%), nearly half (46%) of associates over 50 (who constituted about 7% of Gap's workforce in these stores) used Shift Messenger to post or pick up a shift.

The majority of posted shifts were covered through Shift Messenger, but coverage through the app decreased as the shift approached. Nearly three-fourths (72%) of shifts posted at least a week in advance got picked up on the app – as were half of the shifts posted only a day to three days in advance. The portion dropped to 30% for those shifts posted less than 24 hours in advance. Overnight shifts and unpopular shifts such as weekend closings were less likely to be covered on the app than other shifts.[22] When asked on the employee survey what had happened when a shift you posted on the app was not picked up by a coworker, 71% of respondents said they worked the shift themselves.

Response time to posted shifts was swift. Over two-thirds (68%) of all completed swaps on the app were completed within 24 hours of being posted and almost a quarter (24%) within five minutes.

On average, hours were lost rather than gained. Shift Messenger users averaged one fewer hour than they were originally scheduled during weeks when they used the app; this was not the case in weeks they did not use the app. Since Shift Messenger was available to associates in treatment stores throughout the intervention period, tech-enabled schedule swaps and changes are a likely source of the greater mismatch between scheduled and worked hours during intervention than non-intervention weeks. As reported above, during intervention weeks associates worked on average 20 minutes less than they were originally scheduled, but during non-intervention weeks, associates worked about 5 minutes more than they were originally scheduled.

Overall, respondents in treatment stores reported greater input into the days they work than

respondents in control stores. A significantly higher percentage of respondents from treatment than control stores reported at least some input into the days they work and the days they have off. On the post-intervention survey, 75% of respondents from treatment stores agreed that they have input into the days they worked each week, compared to 66% of control store respondents ($p<.05$); 70% of respondents from treatment stores reported input into days off versus 61% of respondents in control stores ($p<.01$).

Input into the number of weekly hours was relatively rare, but greater in treatment stores. Barely half (53%) of respondents in treatment stores and less than half (42%) of respondents in control stores reported at least some input into the number of hours they work ($p<.01$).

5. A lot of instability remained

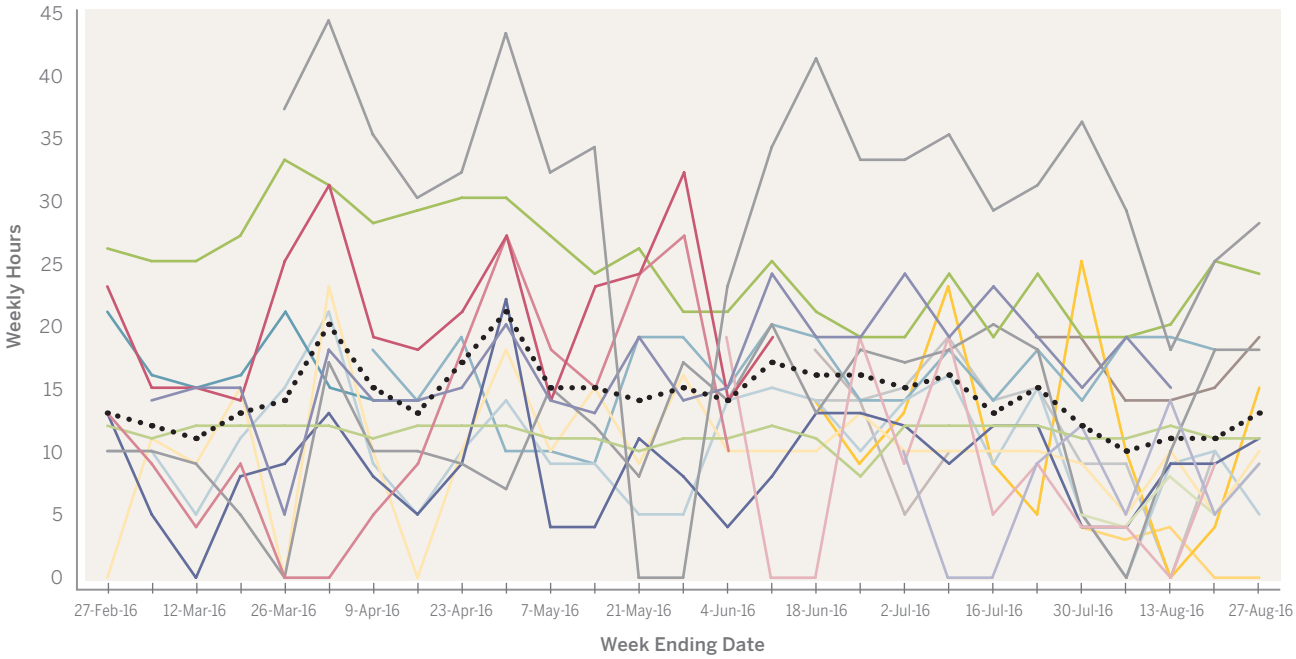
Although the experiment shifted schedules towards stability on three out of four dimensions, a lot of instability remained. Figure 6 illustrates that weekly hours remained quite unstable for part-time associates during the intervention period, even though the average for all associates remains relatively stable. This pattern is typical of stores in our sample and suggests there is room for more stable scheduling than our experiment achieved.

Summary of findings on how stable schedules became

During the Intervention, associates' shift times, start and end times, and number of weekly work hours became more consistent. This greater consistency reflects the goals of our two intervention components of Core Scheduling and Stable Shift Structure. Associates in treatment stores reported greater predictability in number of weekly hours and fewer shifts were changed from the original posted schedule. Associates who were designated as Part-time Plus saw an increase in hours over the intervention period; however, the average associate did not. During intervention weeks, associates averaged about 20 minutes less of work than originally scheduled, and during weeks associates used Shift Messenger, they averaged an hour less of work than originally scheduled. Even though – on average – associates lost rather than gained hours when using Shift Messenger, the majority of associates with access to the app used it to gain additional control over their work schedule, with associates in treatment stores reporting greater input in their schedule than those in control stores. Overall, the different dimensions of schedule stability increased by amounts that seem modest – which makes the substantial business benefits that emerge from the pilot all the more striking.

Figure 6

Scheduled Hours by Week (All part-time associates in a treatment store)



Each solid line represents 1 employee, while the dotted line shows the average for all 21 part-time associates with scheduled hours in the selected period. Note the wide variation in individuals' weekly hours as compared with the more stable average.

Overall, the different dimensions of schedule stability increased by amounts that seem modest - which makes the substantial business benefits that emerge from the pilot all the more striking.

Did More Stable Scheduling Improve Business Outcomes?

Given that this is the first study of its kind, we have no benchmarks that would allow us to tell whether more stability than was delivered through the experiment would produce even larger business benefits. No doubt there's a sweet spot that minimizes both the costs associated with too much variability and the costs associated with too little. We leave it to future research to identify this sweet spot.

That said, our most important finding is that companies may not currently be operating at the stability sweet spot. The experiment suggests that a shift to more stable scheduling can provide highly significant business benefits for many retailers by increasing both sales and labor productivity.

More stable scheduling increased sales

We performed a difference-in-difference analysis to estimate the effect of the Intervention on treatment stores. This technique is widely used in economics, finance, marketing, and many other fields. It allows us to control for exogenous variation in observable factors such as traffic that could vary across our treatment and control stores, even with prior randomization. Controlling for these factors is especially useful as our sample of stores is small and in conjunction with randomization will help to balance differences across treatment and control stores.

Shifting schedules towards greater stability sharply increased median sales. The increase in sales due to our intervention was high – 7% – in an industry in which companies often work hard to achieve increases of 1–2%. Even more remarkable is the fact that the sales increase was achieved without investments in increasing traffic; that is, sales increased due to better conversions and higher basket values.[23] Our findings confirm business school studies' suggestion that too much labor flexibility hurts performance (Kesavan et al., 2014; Ton, 2014). While using too many part-time or temporary employees has been found to lead to lower sales and higher costs, schedule variability has not previously been identified as one of the contributing factors to the lower sales.

Increased sales meant that treatment stores earned an average of $4,363 per week more than control stores. Over the 35 weeks of the pilot, we estimate that shifting to more stable scheduling yielded $2.9 million in increased revenues for Gap in the 19 treatment stores. The only out-of-pocket cost associated with stable scheduling was roughly $31,200 spent on Targeted Additional Staffing, which resulted from adding 1 or 2 sales associates for a 4-hour window, 7 days a week. Taking into account only this direct investment, the ROI on Gap's investment in stable scheduling was high. We did not take into account the time devoted by managers and staff to attend the initial roll-out meeting and any additional time they may have spent to implement the Intervention as we do not have good estimates of those times. We believe these amounts to be small, but without exact numbers, we cannot calculate the exact ROI. If stable scheduling were adopted enterprise-wide, transition costs might well entail the costs of upgrading or replacing existing software systems.

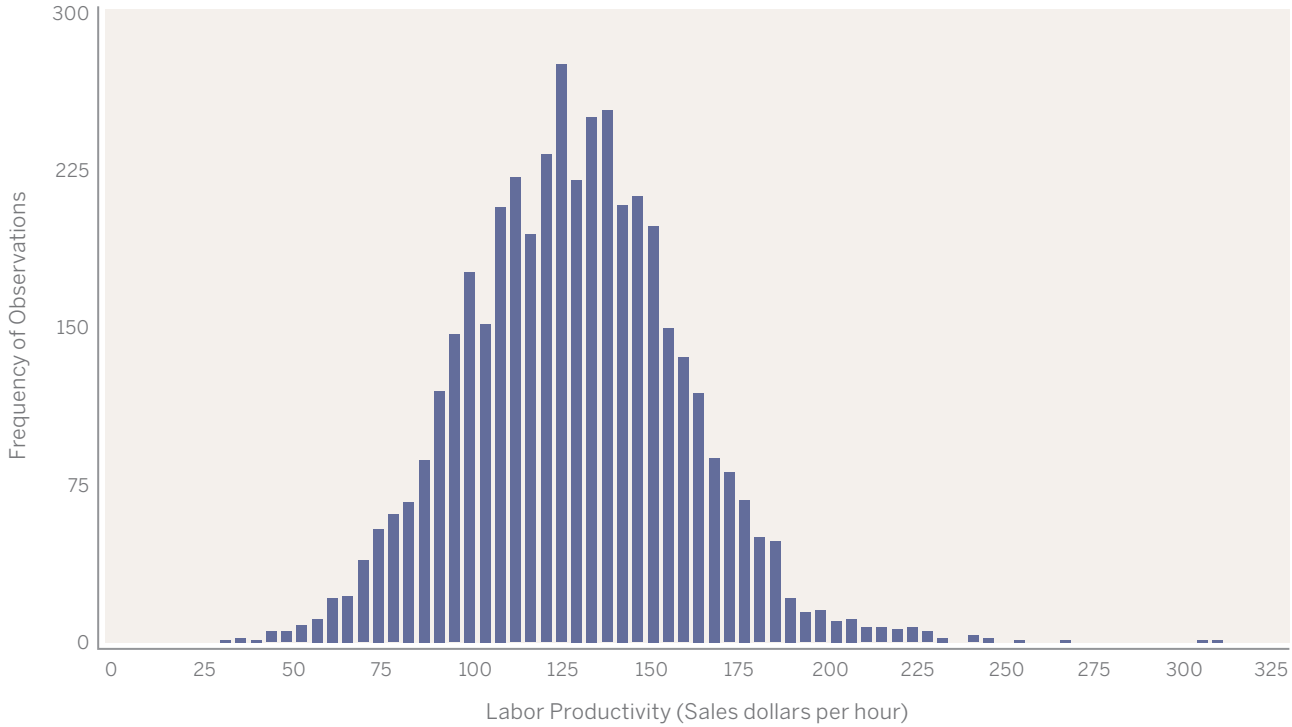
Over the 35 weeks of the pilot, we estimate that shifting to more stable scheduling yielded $2.9 million in increased revenues for Gap.

More stable scheduling increased labor productivity

Given that stable scheduling increased sales, the natural follow-up question is why. The answer is that labor productivity (sales per hour) increased by 5% as a result of our intervention. To put our results in context, retail labor productivity grew by about 2.5% per year between 1987 and 2014 (Hortaçsu & Syverson, 2015). Treatment stores generated an additional $6.20 of revenue per hour of labor than did control stores. Figure 7 represents the distribution of labor productivity across store-weeks for participating Gap stores during the experiment. Mean Productivity = $125 per hour.

Figure 7

Distribution of Productivity ($/hr)



The next question is why productivity increased. One probable factor was that turnover among more experienced staff decreased in treatment stores as a result of our intervention. Preliminary results show that the average tenure among employees who quit in the treatment stores decreased by 10.8 months. The average tenure of associates at the time of quitting is 25.5 months. More experienced staff presumably are more adept at selling, which is why losing them hurts productivity.

Consistent with this turnover pattern, we observe workers with more experience benefitting more from significant reduction in instability in treatment stores compared to less experienced associates. We measured instability using the coefficient of variation (CV) of weekly hours worked by associates (standard deviation in weekly hours / average hours). The average CV in our sample was 0.71. Our results show that employees with average tenure experienced a 4% reduction in CV after the Intervention. This decrease was 6% for employees at the 75th percentile of the distribution with 50 months of tenure. While we did not specifically ask managers to target this population, it appears that managers may have favored more experienced workers when implementing the Intervention. One concern we had was whether such improvements in stability for the experienced workers came from a decrease in stability for the less experienced associates. Our statistical tests rule out this possibility as we do not see any decrease in stability for the less experienced workers as a result of our intervention.

Summary of experimental findings on feasibility and business outcomes

Our experimental analyses support three major conclusions about the feasibility and business benefits of improving schedule stability:

1. A widespread assumption in retail – that current levels of schedule instability are inevitable due to fluctuations in customer demand – is inaccurate.
2. It was possible to shift to more stable schedules, even in today's fast-paced, globalized business environment.
3. Doing so increased sales and improved labor productivity, suggesting a high return on investment.

Our findings confirm that a lot of the instability in retail stems not from associates but from headquarters – even as headquarters is convinced that most instability comes from stores.

What Drives Schedule Instability?

A common assumption is that fluctuations in customer demand make unstable retail schedules unavoidable. Our findings challenge this nigh-universal assumption. According to our analyses of detailed data on traffic and labor, only 30% of the variability in weekly payroll hours is explained by changes in traffic from week to week – the same percentage found in a prior study of a different retailer (Kesavan & Kuhnen, 2017).

Qualitative data from biweekly manager interviews[24] and associate focus groups suggest that some instability stems from associates who are unable to work as scheduled. Sometimes associates call in sick at the last moment, or are unable or unwilling to work for other reasons. People lead complicated lives, and retail sales associates are no exception.

But our findings confirm a growing body of research documenting that a lot of the instability in retail stems not from associates but from headquarters – even as headquarters is convinced that most instability comes from stores (Netessine et al., 2010; Fisher, 2012; Ton, 2014, 2017; Amico, 2017).

> "[H]eadquarters make decisions in silos and rarely consider the effect on employee productivity and customer service."
>
> – Zeynep Ton (2017)

Prior research identifies promotions as a key source of instability. Last-minute changes in the timing of promotions often drive last-minute schedule changes, as described by Professor Ton of MIT:

> Merchandising decides to move a promotion from Friday to Wednesday to stimulate demand. This doesn't seem like a big deal at headquarters. But the store manager must shift dozens of hour of labor from Friday to Wednesday (Ton, 2017).

Wild fluctuations in the number of promotions from week to week are another source of last-minute changes (Ton, 2017). Ton's research also notes the role of shipment schedules in fueling instability (Ton, 2017). Ton also identifies the organizational dynamic that means that the impact of promotions and shipment on schedules is not heard at headquarters: "functions at headquarters make decisions in silos and rarely consider the effect on employee productivity and customer service," she notes, so stores have little opportunity to provide feedback about the effects of last-minute changes on stores' ability to maintain high-quality customer service and execution (Ton, 2017). Our research confirmed these conclusions.

The role of headquarters in driving last-minute changes was revealed in our qualitative data, in particular in our biweekly interviews with store managers. These interviews provided rich insights that allow us to address two key mysteries:

1. If only 30% of schedule instability reflects fluctuations in customer demand, what explains the remaining instability?

Contrary to popularly held belief that schedule instability is the result of variable traffic patterns, we find that decisions made at headquarters contributed a lot to schedule instability of workers.

2. Why does unstable scheduling persist if scheduling for more stability could produce better business outcomes?

We find that schedule instability is the outcome of suboptimal algorithms that ignore many hidden costs associated with unstable schedules.

For maximum impact, a stable scheduling initiative should involve not only HR, but also marketing, supply chain, leadership, operations, and finance.

The role of headquarters: HQ Instability Drivers

Although the experiment showed that schedules shifted towards greater stability, with concrete business benefits, our findings also showed that considerable instability remained. While some schedule instability is associate- and customer-driven, store managers overwhelmingly reported that a lot of schedule instability stemmed from decisions made at HQ, confirming prior research at another retailer that found that store managers often do not feel supported by higher levels of management when they try to provide more stable schedules to associates in their stores (Lambert et al., in review).

The HQ drivers of scheduling instability identified by managers fell into three categories: 1) changes in shipment dates or number of units; 2) changes in promotions; and 3) leadership visits. Managers reported that these changes often came down from HQ *after* they had published their schedules. Note that our intervention was launched out of human resources (HR) and

originally designed to affect only store managers – not marketing and supply chain. So it is understandable that other business departments sometimes were not in sync with Stable Scheduling Study priorities. These findings suggest the need for a major paradigm shift to move away from the widespread assumption that stable scheduling is an HR issue. For maximum impact, a stable scheduling initiative should involve not only HR, but also marketing, supply chain, leadership, operations, and finance.

HQ Instability Driver #1: Last-minute changes in shipment units or dates

Supply chains in retail are complex, often leading to last-minute changes that disrupt workers' schedules. Managers lamented the influence of shipment changes on their ability to provide stable schedules: "Lack of shipment planning is going completely against the effort to get the schedule published two weeks in advance."

Some managers felt that shipment was the key operational factor undermining their ability to provide more stable schedules: "Way more shipments are coming than what is predicted. This is the biggest operational issue we are facing."

Inaccurate numbers. Although managers were usually given notice of how many units shipments would contain, these numbers were often grossly inaccurate: "This morning we came in to find out how many boxes we had. We were left in the dark." Sometimes, they received some notice, but much less than the two weeks they would need to build shipment needs into the schedule: "We usually can only get an accurate number two days before the shipment comes." The result was last-minute schedule changes, as reported by many managers. Said one, "On Monday, our shipment units changed drastically. We had to cut a bunch of hours."

Another manager recounted being left with empty shelves when shipment went from a projected 250 boxes to only 11. Some variation in the number of units is probably unavoidable, but sometimes the variation was extreme.

Inaccurate dates. Although less frequent, sometimes shipment dates changed at the last minute, requiring major overhauls to the schedule. One manager shared, "They give us a plan but then it changes. We plan but then I have to cross out the schedule because all of a sudden shipment is coming early, but then nothing arrives so we need to do it all over again."

"They give us a plan but then it changes. We plan but then I have to cross out the schedule because all of a sudden shipment is coming early, but then nothing arrives so we need to do it all over again."

– Store Manager

Sleep deprivation. Shipment is often processed at night or very early in the morning, which means that abrupt changes can affect employees' sleep. For example, a manager explained, "Friday we added three extra bodies to get shipment done. I had four of us come in at 3 a.m. Our little army was sleepy."

As in prior studies, we found that attempts to reduce instability for associates sometimes came at the expense of store managers themselves (Lambert & Henly, 2010; Henly & Lambert, 2015). It was not uncommon for a manager to report shouldering the burden: "I can't ask [associates] to come in at 2 a.m. so I just did it myself."

HQ Instability Driver #2: Changes in promotions and markdowns

Like most retailers, Gap tries to be fleet of foot, and to use promotions to drive the business. But changes in promotions often require a lot of work: items need to be marked down or moved around, tables set up and taken down, windows changed. Many manager comments, such as those listed below, noted promotions' important role in driving schedule instability:

- *"Promotions changing is the main driver of managers having to make schedule changes."*
- *"It's just been a roller coaster with promo changes. This week alone the window changed three times."*
- *"There were four promo changes for this week."*

Promotional changes undercut Two-week Advance Notice. Several managers noted how promotions changes undercut their ability to give associates two

weeks' notice of schedules as required by company policy. For example, one noted, "This is where the frustration with doing the schedule two weeks out comes from. They still do last-minute promotional changes all the time."

Markdowns and promo changes drive long hours. Although some associates may not get enough hours, scheduling changes due to markdowns or promo changes drive very long hours for others, both associates and managers. One manager noted that, as a result of markdowns that got changed at the last minute to Sunday from Monday, she added four extra people and extended one person's shift – and stayed late herself. Another had to ask staff to come in from 6 a.m. to 10 a.m. on a Sunday, and came in herself on her day off.

- *"[Name] just left now. He's been here for 14 hours because of the windows."*
- *"Promos are changing, things are being cancelled, it's a crazy time. There have been leadership working on their days off, working six days a week. I've been working open to close and can't remember the last time I worked a regular work day."*

HQ Instability Driver #3: Leadership visits

Understandably, top corporate leaders want to get out to see what's going on in the stores and store managers want their stores to look good for their visits. But this creates a dynamic where leadership visits disrupt schedules, especially in large stores where these visits are more frequent. Store managers often expressed frustration, such as this one: "They changed a visit from Monday to Thursday and I only found out on Saturday. I made the schedule two weeks ago for extra help on Sunday, but now I have to change it."

Often schedules changed very substantially in preparation for these visits, as reported in the following manager quotes:

- *"We got four days' notice [for this visit]. I had to add in 100–150 hours."*
- *"Probably extended 2–3 shifts every day in the run up to the visit."*

Managers often felt that hours spent on leadership visits could be better spent on the sales floor. One manager recalled using 70 hours to prepare for a visit that never happened. Noted another manager tartly, "If payroll is such a commodity, I feel it's crucial to spend it on the business versus visits. This isn't a museum. You know we aren't curating a museum." The great deal of preparation and prioritizing that stores put into leadership visits may also ultimately perpetuate the false image to corporate leaders that they are able to successfully operate their stores with minimum labor hours.

To our knowledge, this is the first study to document in detail the role of not only shipment and promotional changes, but also of leadership visits, in creating last-minute schedule changes. We hope that this report sheds light on these important drivers on schedule instability and motivates researchers and retailers to identify ways to handle them.

Managers often felt that hours spent on leadership visits could be better spent on the sales floor.

The important message for retailers is that an attempt to move towards schedule stability is a major change management initiative that needs to involve several different departments. Human resources will of course be involved, but improving schedules is not just an HR or operations issue; norms and practices also may need to change around shipment, marketing, leadership visits, labor budgets, and the metrics used to evaluate manager and store performance.

One approach is to use the package of practices that Professor Ton of MIT has christened the "Good Jobs Strategy." A growing body of evidence documents the promise of this approach, and provides a roadmap for implementing it (Ton, 2014, 2017). Yet some retailers may find it challenging to implement all aspects of that strategy. Notably, the retailers Ton highlights (like Costco) have largely or completely eliminated promotions and stock a strictly limited range of products (e.g. Costco's two brands of diapers vs. a grocery store's two shelves of diapers). Our study raises an intriguing question: might retailers be able to offer more stable schedules through less radical changes to supply chain and marketing? This is an empirical question retailers will want to explore. Our results provide evidence that some increase in schedule stability is possible even without changing these other processes.

The role of technology: Suboptimal scheduling algorithms omit instability's hidden costs

The first step in understanding how scheduling algorithms are suboptimal is to understand how schedules are constructed. To construct a schedule of, say, week 20 of this year, a company's finance department looks at sales or traffic in the same store in week 20 of the previous year (Netessine et al., 2010). That number is then adjusted using a proprietary algorithm that includes such factors as trends in the business and the weather. Once the algorithm yields the expected sales (or traffic) in a given store, the finance or operations department translates expected sales (or traffic) into labor hours required, again using a proprietary algorithm. The labor hours are then allocated across workers in a store, prioritizing full-time workers first and then giving the residual hours to

part-time workers based on their availability. Usually the availability of part-time workers is more than the hours they receive so the algorithm distributes the sparse hours across many workers. What is missing in these algorithms are the hidden costs of unstable scheduling.

Our study team spent a lot of time talking to managers in Gap stores as well as people at Gap headquarters. This gave us valuable insights into not only the hidden costs of unstable scheduling but how the move to more stable scheduling reduced them.

Poor execution is a hidden cost of unstable scheduling

Poor execution in retail includes both poor customer service and low "conformance quality" – how well employees execute prescribed processes such as replenishing inventory (Ton & Huckman, 2008; Kesavan & Mani, 2015). A large literature documents the prevalence of poor execution in retail (Salmon, 1989; Raman et al., 2001a, 2001b; DeHoratius & Raman, 2003; Ton & Raman, 2010; Corsten & Gruen, 2003; Ton & Huckman, 2008; van Donselaar et al., 2006; Fisher et al., 2006; Netessine et al., 2010; Fisher et al., 2017). Low conformance quality means that inventory is not replenished, so customers may fail to find what they want, or are misinformed about a product's availability ("phantom stockouts") (Raman et al., 2001a; Kesavan & Mani, 2015); and theft occurs because too few associates are in stores, or because store associates are poorly trained or under-motivated (Ton, 2014). Low service quality means that a customer may not find a sales associate to help find a size, or leaves because the check-out line is too long (Fisher et al., 2006). While prior studies link poor execution with lean staffing, our qualitative data link poor execution with unstable scheduling.

Managers gave clear examples of how they too grappled with these challenges and how our intervention improved execution:

A better match between people and required skills. Computerized scheduling systems treat workers as interchangeable, making it impossible to take things like skillset into account when developing the schedule. With a shift towards stable scheduling, managers felt they were better able to "get the people with the right skills at the right times," to quote one. Other managers added:

- *"Consistency is good for business and customers. You know what you're going to expect from people. They're good at what they do."*
- *"[Name] rings a lot so he's good at knowing what's going on in terms of what code is still good, or expired, or what. He's able to have real dialogue with customers rather than answering questions with, 'I'm not sure.'"*

Improvements in customer service. Associates who worked consistent shifts, and more shifts, could respond effectively to recurring customer needs because they knew more about the store's merchandise. For example, one manager praised the effectiveness of a consistent associate, "[Because she's on a Core Schedule] she's here so often she knows where the product is, and she's able to tell you if we have something or don't."

Managers also report that they felt customer service improved when they could have both a cashier and someone to help customers. "I've definitely seen a customer experience/satisfaction increase, because there is someone on the floor and someone able to ring."

Improved customer service could lead not only to sales, which retailers track closely, but also to long-term loyalty – which often goes unmeasured.

Associates who worked consistent shifts, and more shifts, could respond effectively to recurring customer needs because they knew more about the store's merchandise.

Fewer "phantom stockouts." Operations research has long documented "phantom stockouts" – where customers are incorrectly informed that an item is out of stock because an item is misplaced somewhere in the store (Ton, 2014). A recent study shows that 35% of items left behind in the fitting room of a department store experience phantom stockouts and sales increase significantly with proper recovery operations (Kesavan et al., 2017). Similarly, in our study, managers noted how organizational improvements were related to improved customer service: "Our fitting rooms look a lot better, our floor is better organized, our [customer experience] score is over 80."

Store managers often noted that, when associates were pulled off the floor to cope with an unexpected shipment

or promotions change, refolding and returning items to the shelves ("replenishment") would get postponed:

- *"Even if you change from a 20% to a 30% promo [which is an easier promo change, you] still need to reprint all the signage and change all the signage in the store.... everything has to change in 30 minutes. And it's not like your payroll changes with it... And that needs to be done before store opens, so then your replenishment goes on the back burner."*

Improvements in on-time arrivals. Inconsistent shifts that have different start times can confuse associates and cause more late arrivals. An unexpected result of scheduling more consistent shifts and start times was that managers felt associates were more frequently on time for their shifts because, for example, many now took the same bus to work every workday.

- *"It helps with stability for them because they can remember what time to leave to get here on time with public transportation."*
- *"Knowing that shift is going to be the same week to week. They know what time to leave, what buses to take."*

Less theft. Studies have noted that lean staffing promotes theft (Ton, 2014). Managers from stores that received Targeted Additional Staffing reported feeling that theft had decreased because there were more people on the floor. "It's obvious when we don't have anyone in Kids; that is when they come in to steal. Now with someone there, they'll come in and walk right out."

High turnover and low engagement are hidden costs of unstable scheduling

High turnover is one reason for poor execution. Many studies have documented high turnover in industries that use unstable schedules (Lambert & Waxman, 2005). Operations research in a department store found that unstable schedules lead to higher turnover rate not just among low performers but also among highly productive employees (Kesavan & Kuhnen, 2017). Kesavan and Kuhnen (2017) estimate that the turnover costs for the retailer in their study may be as high as 20% of total payroll cost for the chain.

The pilot did not decrease overall turnover in treatment stores. This may be because the Intervention was in place too short a time to do so. The pilot did decrease turnover among more experienced store associates, who arguably are most valuable to their employers. In some stores, turnover was so high that the store manager was always hiring. "We don't have a lot of people who can work some of the shifts. So we try to hire for that, but it's like a leaking bucket. You hire one person and another person quits." Other managers noted that short, part-time hours fuel turnover. "New hires ask for more hours and ask to be trained, but it never really happens. So they end up getting another job."

Qualitative data show various ways turnover costs are hidden. Payroll budgets often failed to include all the costs of on-boarding and training. Managers reported that they usually received between 8 – 12 hours to onboard a new hire. All felt more time was needed. "We don't get funded for appropriate training. We can only train them on the very basics and after that they are expected to be fully functional on the floor." Other managers linked high turnover with execution problems that resulted from insufficient training.

- *"When new hires don't get trained – that's when the floor looks bad."*
- *"Sometimes we get new merchandise in and the new hires accidentally mark them down because they don't know the difference."*

- *"At my store, someone left a sensor on one of the purchases. They don't teach them how to do things."*

Even if employees do not leave due to unstable scheduling, they may become demoralized and disengaged. "When you schedule willy-nilly all the time people just come in and do the bare minimum to get paid and then leave. Just enough to not get fired." But with stable scheduling, another manager noted a marked change: "Once they have the stability every week, they get more invested emotionally and to drive the business."

Though *quantitative* data found no overall decrease in turnover, *qualitative* data found managers expressing the view that giving associates more schedule stability reduced turnover. "Everybody likes routine in life, and they deserve it. Monetarily it's hard for us to give people raises but what we can do is to get people the schedules they want and need."

The hidden costs of managing complex schedules

Computerized scheduling systems were marketed with the promise of saving managers time on scheduling (Pillar, 2006), with cheerful promises that computerized systems "can boost productivity by freeing up managers" (Maher, 2007). But that was on the assumption that the computer generated schedules were good to go without manager intervention. In fact, managers found they needed to devote time to edit the schedules generated by Gap's computerized system, which often needed extensive work. Prior research at a different retailer found that 70% of schedules were edited by store managers (Bernstein et al., 2017).

Once we pushed managers to stop relying so heavily on computer-generated schedules, and to try to schedule in more week-to-week consistency, they found that scheduling took less time. "I'm excited that the schedule is so stable. It takes less time to write; it used to take 3–4 hours. Now it takes an hour and a half," said the manager of a medium-sized store. In large stores, the savings could be dramatic: "When I got to this location, they were spending three days doing the schedule, now I can do it in four hours," noted the manager of a large store. The two practices that managers said helped them the most in decreasing the time they spent on the schedule were Stable Shift Structure (stable starting and ending times) and Core Scheduling (giving more associates a core schedule of shifts they could count on from week to week).

> "When you schedule willy-nilly all the time people just come in and do the bare minimum to get paid and then leave. Just enough to not get fired."
>
> – Store Manager

Because the time managers spend on scheduling typically is not tracked, no feedback mechanism existed by which HQ could recognize the way managing unnecessarily complex schedules wasted managers' time. More stable scheduling allowed managers to spend less time on the schedule and more on other aspects of the business more directly linked with increasing sales.

Improved schedule stability not only took managers less time; it also relieved pressures on managers in other ways. Managers found that they could begin to rely more on their Part-time Plus staff, who could operate more independently and needed less supervision. One manager reported, "With the two Part-time Plus people, they don't have to constantly be coached. They just come in and start their day and don't have to ask questions to leadership." Management transitions became easier, too. "It's good to have people [who] are consistent, especially right now with no store manager. They know exactly what they should be doing." So did manager vacations. "I was very hesitant about this component initially, but when I was gone for two weeks one time, I had my schedule managers just use the core schedule and build from there and he built around it."

The hidden costs of a low-availability workforce

To our knowledge, no prior research has highlighted the costs associated with hiring a low-availability workforce. Our data pinpoint the hidden costs of this labor strategy.

Because managers could offer only short part-time hours, some employees had to get a second or third job, thereby limiting their availability to Gap; according to our employee survey, 49% of part-time Gap associates held a second job. This meant that store managers found themselves competing with their employees' other employers. "I asked associates to stay later. A

couple said no: they either had to go to their second jobs, childcare, or something else."

Spring break, mid-terms, summers, new semester – all can spell disruption of businesses that rely on a low-availability workforce, because workers who want a low-hours schedule often are students or have other responsibilities. But an organization dependent on students means business disruptions of various sorts. Store managers persistently noted that school vacations, changes in class schedules, or end of semester assignments could disrupt normal store operations.

Once a company gets the reputation of offering only short part-time hours, the only workforce they can attract is a low-availability workforce. Thus some store managers who wanted to give more employees Part-time Plus found they could not do so because their associates only wanted the short hours they were already working. "It's not that people don't want to take more shifts; they're just not available. We need to hire more people with more availability." This creates a vicious-cycle for associates as well, who may want to work more hours – as about 50% of Gap part-time associates reported. Since they cannot count on additional hours, associates load up their non-work time with other responsibilities and then cannot accept additional hours when they pop up.

The costs of a low-availability workforce also are not typically counted as costs of unstable schedules – but these costs are very real. Perhaps for that reason, Gap has now moved towards offering additional hours to people already on staff before hiring additional staff (J. Blazek, personal communication, October 26, 2017).

Unstable scheduling looks good because its true costs are hidden

Store managers, from our very first interviews, expressed frustration that HQ gave them too few hours to run their stores for maximum sales and productivity. Our study helped explain why costs clearly visible in the stores remained invisible at headquarters. The power differential between HQ and stores meant that headquarters did not receive this information, or did not process it. As a result, very substantial costs of unstable scheduling were hidden, including the real cost of labor, the high costs of poor execution, the amount of time spent managing complex schedules, and the costs of a low-availability workforce. With all these costs deep-

sized, no wonder unstable scheduling looks attractive. It's easy to justify a business practice if you count its benefits but not its costs (Williams & Huang, 2011).

If the hidden costs are visible to store managers, why are they not incorporated into staffing algorithms?

Computerized scheduling software has been marketed as shifting from intuitive to scientific scheduling. To quote the website of one of the large purveyors of scheduling software, "Kronos labor forecasting and scheduling solutions take the guesswork out of aligning the right talent, in the right place, at the right time. Our automated solutions eliminate manual scheduling processes to help ensure optimal labor coverage for every shift, every day" (Kronos, 2017). Another vendor noted, "Our workforce scheduling solutions use data from your key business systems to create demand-based schedules avoiding any over or understaffing" (Workforce Software, 2017). At the start of the Stable Scheduling Study, HQ seemed to take that purveyors' claim to offer scientific scheduling at face value. We heard from store managers that they were discouraged from changing the schedules generated by "PM" – Gap's computerized scheduling.

Some store managers felt that HQ had an unrealistic assessment of PM's schedules: "[HQ] thinks the schedule is perfect when it gets spit out of PM, but it's so far from the truth," noted one. In contrast, some at HQ attributed managers' dissatisfaction with PM to "emotional scheduling" – a script pushed by the purveyors of scheduling software. "Emotional feedback from the field must be vetted through sampling and analysis, and retailers must guard against making knee-jerk changes to the labor model based on anecdotal feedback," advises a paper on best scheduling practices from Kronos (Strohecker, 2017). "HQ still trusts PM not managers," read our early study notes. Our impression was that this attitude shifted during the course of the Stable Scheduling Study. By the end of the study, we heard less about emotional scheduling.

The charge of "emotional scheduling" is a classic example of what organizational behavior researchers call "agency misattribution." Agency misattribution occurs when one unit of a company attributes coordination failures to the personality drawbacks of colleagues in another unit, instead of recognizing that the other unit simply possesses different information. Agency misattribution is a common reason miscommunication arises, and communication failures persist, in organizations (Heath & Staudenmayer, 2000), and may have played a role in making it difficult for information about the hidden costs of stable scheduling to flow from stores to headquarters. We confirmed the finding of prior researchers (Netessine et al., 2010; Fisher, 2012; Ton, 2014, 2017; Amico, 2017) that the power dynamic between HQ and stores impedes communication around labor planning in retail environments:

> Typically, labor planning is done at retail headquarters, with a lot of thought but little input from store managers going into it. Then the plan is given to the store manager, who must do whatever is necessary to comply (Netessine et al., 2010, p. 2).

In short, companies' commitment to unstable scheduling reflects not business realities but an organizational failure.

If poor communication is one factor that results in suboptimal labor scheduling algorithms, equally or more important is the misalignment of incentives. Because labor expenses account for such a large percentage of controllable costs, HQ may be tempted to cut labor hours to "make their numbers" for the next quarterly report (Fisher, 2012; Kesavan et al., 2014). Even if the cut in hours makes it difficult or impossible for store managers to optimize store performance, HQ makes its numbers and stores are left to cope the best they can.

In short, companies' commitment to unstable scheduling reflects not business realities but an organizational failure. HQ's incentive is to "make its numbers" in the short term, which it can do most easily by cutting labor hours through practices that fuel schedule instability. This depresses sales and labor productivity in the long term, but that information gets lost when last year's (artificially depressed) sales numbers are used to calculate this year's labor hours. The result is sales and productivity chronically lower than they could be: store managers know it, but the information never gets through to headquarters.

It's easy to justify a business practice if you count its benefits but not its costs.

Conclusion

This study contests the widespread conventional wisdom that unstable schedules are impossible and/or uneconomic in today's fast-paced, low-margin retail environment. In a randomized controlled experiment, we found that giving employees more schedule stability actually increased both sales and labor productivity, signaling a high return on investment. More stables schedules are a win-win for both retailers and their employees. This finding may change the social dynamics surrounding scheduling practices. Thus far, unions and other workers' rights advocates have been the chief proponents of more stable scheduling for hourly workers. Our findings suggest that, in the future, the movement towards more stable scheduling may be led by stockholders.

A second major contribution of this study is to provide new insight into the steps companies need to take in order to move towards more stable scheduling. While the experiment shifted associates toward greater schedule stability, considerable instability remained. We confirmed key findings of Professor Ton of MIT, who also documents how last-minute changes in shipments and promotions fuel schedule instability (Ton, 2014, 2017). As a solution to these and other challenges, she advocates the "Good Jobs Strategy" based on her case studies of companies like Costco that offer both good jobs and good profits. Yet the Good Jobs Strategy abandons practices that are central to the business model of many retailers by eliminating promotions altogether and offering a strictly limited range of merchandise. Our study raises an intriguing question: might retailers be able to offer more schedule stability than our experiment produced through less radical changes to supply chain and marketing? This is an empirical question retailers will want to investigate.

> More schedule stability actually increased both sales and labor productivity, signaling a high return on investment.

Particularly in view of the business benefits, employers should be encouraged to take the leap to more stable scheduling in ways that work for their business, given that several cities (San Francisco, Seattle, New York City, and Emeryville, California) and two states (Oregon and New York) have already passed scheduling legislation and comparable legislation is pending in at least 13 additional municipalities. Laws vary, but typically legislation sets minimum advance notice and requires employers to compensate employees (with "predictability pay") for last-minute schedule changes.

Simultaneously, states' attorneys general are taking legal steps to pressure employers to eliminate problematic scheduling practices. A 2015 inquiry by New York State Attorney General Eric Schneiderman alleged that the use of on-call shifts violates state "reporting pay" regulations. Subsequently, six national retail brands ended on-calls. In 2016, eight additional attorneys general joined Schneiderman to apply pressure on 15 national retailers (New York State Office of Attorney General, 2016). Since that time all 15 have either discontinued on-calls or committed to doing so nationwide. Continued pressure from legislatures and lawsuits means that the time is right for employers to take the initiative to improve schedule stability in ways that work for them.

More stable schedules are a win-win for both retailers and their employees.

WORKS REFERENCED

Amico, L. (2017). Clocking In: What It's Like to Work a Bad Job. December 2017. *Harvard Business Review*. Retrieved from https://hbr.org/2017/12/clocking-in-what-its-like-to-work-a-bad-job.

Bailyn, L., Collins, R., & Song, Y. (2007). Self-Scheduling For Hospital Nurses: An Attempt and Its Difficulties. *Journal of Nursing Management, 15*, 72–77.

Bernstein, E., Kesavan, S., Staats, B.R., & Hassall, L. (2014). Belk: Towards Exceptional Scheduling. *Harvard Business Review*. Case Study 415–023, September 2014. (Revised February 2017.)

Bloom, H.S. (2005). *Learning More from Social Experiments*. New York: Russell Sage Foundation.

Boushey, H. & Ansel, B. (2016). Working By the Hour: The Economic Consequences of Unpredictable Scheduling Practices. Washington Center for Equitable Growth. Retrieved from http://equitablegrowth.org/research-analysis/working-by-the-hour-the-economic-consequences-of-unpredictable-scheduling-practices.

Bureau of Labor Statistics. (2017). Table 39: Median Weekly Earnings of Full-time Wage and Salary Workers by Detailed Occupation and Sex. Labor Force Statistics from the Current Population Survey. Retrieved from https://www.bls.gov/cps/cpsaat39.htm.

Carré, F. & Tilly, C. (2017). *Where Bad Jobs are Better: Retail Jobs across Countries and Companies*. New York: Russell Sage.

City and County of San Francisco. (2016). Formula Retail Employee Rights Ordinances. Retrieved from http://sfgov.org/olse/formula-retail-employee-rights-ordinances.

Clawson, D. & Gerstel, N. (2014). *Unequal Time: Gender, Class, and Family in Employment Schedules*. New York: Russell Sage Foundation.

Corser, M. (2017). Job Quality and Economic Opportunity in Retail: Key Findings from a National Survey of the Retail Workforce. Center for Popular Democracy. Fair Workweek Initiative. Available at https://populardemocracy.org/news/publications/job-quality-and-economic-opportunity-retail-key-findings-national-survey-retail.

Corsten, D. & Gruen, T. (2003). Seeking On-Shelf Availability – An Examination of the Extent, the Causes and the Efforts to Address Retail Out-Of-Stocks. *International Journal of Retail and Distribution Management, 31*(12), 605–716.

DeHoratius, N. & Raman, A. (2003). Building on Foundations of Sand? *ECR Journal, 3*(1), 62–63.

Fisher, M.L. (2012). Retail Rage. *Harvard Business Review*. January 6, 2012. Retrieved from https://hbr.org/2012/01/retail-rage.

Fisher, M.L., Gallino, S., & Netessine, S. (2017). Setting Retail Staffing Levels: A Methodology Validated with Implementation. INSEAD Working Paper No. 2017/36/TOM. Retrieved from http://pulsar.wharton.upenn.edu/a8d0f9d/settingretailstaffinglevels.pdf.

Fisher, M.L., Krishnan, J., & Netessine, S. (2006). Retail Store Execution: An Empirical Study. The Wharton School, University of Pennsylvania, Philadelphia, PA. Retrieved from https://faculty.wharton.upenn.edu/wp-content/uploads/2012/04/FisherKrishnanNetessine.pdf.

Forbes. (2017). Gap. Retrieved from https://www.forbes.com/companies/gap. Accessed February 21, 2018.

Golden, L. (2016). Still Falling Short on Hours and Pay: Part-Time Work Becoming New Normal. Economic Policy Institute. Retrieved from http://www.epi.org/publication/still-falling-short-on-hours-and-pay-part-time-work-becoming-new-normal.

Hammer, L.B., Johnson, R.C., Crain, T.L., Bodner, T., Kossek, E.E., Davis, K.D., Kelly, E.L., Buxton, O.M., Karuntzos, G., Chosewood, L.C., & Berkman, L. (2016). Intervention Effects on Safety Compliance and Citizenship Behaviors: Evidence from the Work, Family, and Health Study. *Journal of Applied Psychology, 101*(2), 190–208.

Hammer, L.B., Kossek, E.E., Anger, W.K., Bodner, T., & Zimmerman, K. (2011). Clarifying Work-Family Intervention Processes: The Roles of Work-Family Conflict and Family-Supportive Supervisor Behaviors. *Journal of Applied Psychology, 96*(1), 134–150.

Heath, C. & Staudenmayer, N. (2000). Coordination Neglect: How Lay Theories of Organizing Complicate Coordination in Organizations. *Research in Organizational Behavior, 22*, 153–191.

Henly, J.R. & Lambert, S.J. (2014). Unpredictable Work Timing in Retail Jobs: Implications for Employee Work-Life Outcomes. *Industrial and Labor Relations Review, 67*(3), 986–1016.

Henly, J.R. & Lambert, S.J. (2015). Work Scheduling Study: A Profile of Retail Sales Associates in a Women's Apparel Firm. University of Chicago, School of Social Service Administration. Retrieved from https://ssascholars.uchicago.edu/sites/default/files/work-scheduling-study/files/wss_profile_of_retail_sales_associates.pdf.

Hortaçsu, A. & Syverson, C. (2015). The Ongoing Evolution of US Retail: A Format Tug-Of-War. *Journal of Economic Perspectives, 29*(4), 89–112.

Keppel, G. (1991). *Design and Analysis: A Researcher's Handbook* (3rd edition). Englewood Cliffs, NH: Prentice-Hall.

Kesavan, S. & Kuhnen, C.M. (2017). Demand Fluctuations, Precarious Incomes, and Employee Turnover. Submitted to the *Journal of Financial Economics*. Retrieved from http://public.kenan-flagler.unc.edu/faculty/kuhnenc/research/kesavan_kuhnen.pdf.

Kesavan, S., Lee, H., & Deshpande, V. (2017). Increasing Sales by Managing Congestion in Self-Service Environments: Evidence from a Field Experiment. Retrieved from http://public.kenan-flagler.unc.edu/faculty/kesavans/fittingroom.pdf.

Kesavan, S. & Mani, V. (2015). An Overview of Industry Practice and Empirical Research in Retail Workforce Management. In: Agrawal N., Smith S. (eds) *Retail Supply Chain Management. International Series in Operations Research & Management Science,* vol 223. Springer, Boston, MA.

Kesavan S., Staats, B.R., & Gilland, W. (2014). Volume Flexibility in Services: The Costs and Benefits of Flexible Labor Resources. *Management Science, 60*(8), 1884–1906.

Kronos webpage. (2017). https://www.kronos.com.

Lambert, S.J., Fugiel, P.J., & Henly, J.R. (2014). Precarious Work Schedules among Early Career Employees in the U.S.: A National Snapshot. Research brief issued by EINet (Employment Instability, Family Well-being, and Social Policy Network) at the University of Chicago. Retrieved from https://ssascholars.uchicago.edu/sites/default/files/work-scheduling-study/files/lambert.fugiel.henly_.precarious_work_schedules.august2014_0.pdf.

Lambert, S.J., Haley-Lock, A., & Henly, J.R. (2012). Schedule Flexibility in Hourly Jobs: Unanticipated Consequences and Promising Directions. *Community, Work and Family, 15*(3), 293–315.

Lambert, S.J. & Henly, J.R. (2010). Work Scheduling Study: Managers' Strategies for Balancing Business Requirements with Employee Needs. University of Chicago, School of Social Service Administration. Retrieved from https://ssascholars.uchicago.edu/sites/default/files/work-scheduling-study/files/univ_of_chicago_work_scheduling_manager_report_6_25_0.pdf.

Lambert, S.J. & Henly, J.R. (2012). Frontline Managers Matter: Labour Flexibility Practices and Sustained Employment in Hourly Retail Jobs in the U.S. In *Are Bad Jobs Inevitable? Trends, Determinants and Responses to Job Quality in the Twenty-First Century.* 143–59. England: Palgrave Macmillan.

Lambert, S.J. & Henly, J.R. (2014). Measuring Precarious Work Schedules. Working paper of EINet (Employment Instability, Family Well-being, and Social Policy Network). Retrieved from https://ssascholars.uchicago.edu/sites/default/files/einet/files/managingprecariousworkschedules_11.11.2015.pdf.

Lambert, S.J., Henly, J.R., Schoeny, M., & Jarpe, M. (in review). Increasing Schedule Predictability in Hourly Jobs: Results from a Randomized Experiment in a US Retail Firm. Available from author.

Lambert, S.J. & Waxman, E. (2005). Organizational Stratification: Distributing Opportunities for Balancing Work and Personal Life. In Kossek, E. & Lambert, S.J., Eds. (2005) *Work and Life Integration: Organizational, Cultural, and Individual Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.

Maher, K. (2007). Wal-Mart Seeks Flexibility in Worker Shifts. *The Wall Street Journal*, 3 Jan. 2007, 1. Retrieved from https://www.wsj.com/articles/SB116779472314165646.

Mandel, M. (2017). How Ecommerce Creates Jobs and Reduces Income Inequality. Progressive Policy Institute. Retrieved from http://www.progressivepolicy.org/wp-content/uploads/2017/09/PPI_ECommerceInequality-final.pdf.

Mani, V., Kesavan, S., & Swaminathan, J.M. (2014). Estimating the Impact of Understaffing on Lost Sales and Profitability in Retail Stores. *Production and Operations Management, 24*(2), 201–218.

McCrate, E. (2017). Unstable and On-Call Work Schedules in the United States and Canada. International Labour Office, Conditions of work and employment series, no. 99: http://www.ilo.org/travail/info/working/WCMS_619044/lang--en/index.htm.

Netessine, S., Fisher, M.L., & Krishnan, J. (2010). Labor Planning, Execution, and Retail Store Performance: An Exploratory Investigation. The Wharton School, University of Pennsylvania.

New York State Office of the Attorney General. (2016). A.G. Schneiderman and Eight Other State Attorneys General Probe Retailers Over Use Of On-Call Shifts. [Press Release]. Retrieved from https://ag.ny.gov/press-release/ag-schneiderman-and-eight-other-state-attorneys-general-probe-retailers-over-use-call.

Perdikaki, O., Kesavan, S., & Swaminathan, J.M. (2012). Effect of Traffic on Sales and Conversion Rates of Retail Stores. *Manufacturing and Service Operations Management, 14*(1), 145–162.

Pillar, M. (2006). A Workforce Optimized. Integrated Solutions for Retailers. Kronos.

Raman, A., DeHoratius, N. & Ton, Z. (2001a). The Achilles Heel of Supply Chain Management. *Harvard Business Review*, May, 136–152. Retrieved from https://hbr.org/2001/05/the-achilles-heel-of-supply-chain-management.

Raman, A., DeHoratius, N., & Ton, Z. (2001b). Execution: The Missing Link in Retail Operations. *California Management Review, 43*(3), 136–152.

Salmon, W.J. (1989). Retailing in the Age of Execution. *Journal of Retailing, 65*(3), 368–378.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. New York: Houghton Mifflin.

Strohecker, S. (2017). Best Practices for Measuring the Customer Experience and Success of the Labor Model, Part 3 of a 3-Part Series: Measuring the Success of a Retail Labor Model. Kronos Inc. Retrieved from https://www.kronos.com/resources/best-practices-measuring-success-retail-labor-model.

Ton, Z. (2009). The Effect of Labor on Profitability: The Role of Quality. Working paper. Harvard Business School. Retrieved from http://www.hbs.edu/faculty/Publication%20Files/09-040_146640ac-c502-4c2a-9e97-f8370c7c6903.pdf.

Ton, Z. (2014). *The Good Jobs Strategy: How the Smartest Companies Invest in Employees to Lower Costs and Boost Profits*. Boston, MA: New Harvest.

Ton, Z. (2017). The Case for Good Jobs. November 2017. *Harvard Business Review*. Retrieved from https://hbr.org/cover-story/2017/11/the-case-for-good-jobs.

Ton, Z. & Raman, A. (2010). The Effect of Product Variety and Inventory Levels on Retail Store Sales: A Longitudinal Study. *Production and Operations Management, 19*(5), 546–560.

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

van Donselaar, K.H., Gaur, V., van Woensel, T., Broekmeulen R.A.C.M., & Fransoo, J.C. (2006). An Empirical Study of Ordering Behavior of Retail Stores. Working paper, Eindhoven University of Technology. Retrieved from https://pdfs.semanticscholar.org/560d/7f414cc82facd1f64c7b79641915369c4b62.pdf.

Walton, S. & Huey, J. (1993). *Sam Walton: Made in America*. New York: Bantam.

Williams, J.C. & Boushey, H. (2010). The Three Faces of Work-Family Conflict: The Poor, the Professionals, and the Missing Middle. Center for American Progress. Center for WorkLife Law. Retrieved from https://www.americanprogress.org/issues/economy/reports/2010/01/25/7194/the-three-faces-of-work-family-conflict.

Williams, J.C. & Huang, P. (2011). Improving Work-Life Fit in Hourly Jobs: An Underutilized Cost-Cutting Strategy in a Globalized World. Center for WorkLife Law. Retrieved from http://worklifelaw.org/pubs/ImprovingWork-LifeFit.pdf.

Workforce Software webpage. (2017). https://www.workforcesoftware.com.

NOTES

1 Notable prior intervention studies include Lambert et al.'s (in review) randomized study of an intervention designed to improve schedule predictability in hourly retail jobs, Bailyn et al.'s (2007) pre/post intervention study of self-scheduling among nurses in a hospital unit, and a set of randomized studies evaluating an intervention to increase supervisor support for work-family issues in both grocery stores (Hammer et al., 2011) and healthcare settings (Hammer et al., 2016).

2 Stores covered by San Francisco's Retail Workers Bill of Rights were excluded from the roll-back.

3 These statistics come from analyses of items in the 2016 General Social Survey-Fluctuating Work Hours Module, developed by Susan Lambert and Julia Henly, University of Chicago.

4 We followed suggestions to match groups (stores) prior to random assignment (Bloom, 2005; Shadish et al., 2002) to minimize the possibility of observed baseline differences between the treatment and control stores, eliminating variance due to the matching variable from the overall error variance (Keppel, 1991). Participating stores in San Francisco (and then separately for Chicago) were rank ordered in terms of headcount (number of employees on the payroll), yearly sales, and last month's sales. Because sales are a primary driver of the number of hours managers have to staff their store and headcount is a common marker of store size and capacity, both are factors that play a role in managers' scheduling practices and baseline differences between control and treatment stores could confound potential intervention effects.

5 Although this additional attention in treatment stores could bias both managers' and employees' accounts of their experiences with scheduling, research staff maintained biweekly contact with managers in both control and treatment stores, and many of our analyses are based on firm data which are not subject to response bias.

6 In future analyses, we will consider the joint contributions of both the intervention and the legislation to schedule stability, to the extent permitted by the small number of stores (3 treatment, 2 control) in our experiment covered by the legislation.

7 We also secured biweekly payroll reports, useful for examining fluctuations in earnings and clarifying the daily worked hour records.

8 When making comparisons between employees in treatment and control stores, we narrow the samples to baseline and intervention periods. That is, baseline responses are from surveys either completed before the all-store meeting at the respondent's store, at which time the Intervention was rolled out (N=337; N=306 part-time only), or by December 31, 2015, whichever came first. Intervention period responses include all respondents who participated in the wave 2 survey (N=478; N=413 part-time only). We describe sources of non-response bias in a subsequent section.

9 Meeting the store hours' budget was a factor in manager performance evaluations. Near the end of the pilot, Gap changed its performance evaluation system in a way that gave more weight to payroll compliance, pushing managers to be extra vigilant about not running over hours.

10 In these analyses, non-intervention weeks include weeks in control stores plus weeks in treatment stores prior to the implementation of the Intervention. Intervention weeks include weeks in treatment stores subsequent to the adoption of the Intervention.

11 For these comparisons, quarter refers to quarters of the annual year. For most measures, schedule stability peaks during 2016-Quarter 2 and then tapers off into 2016-Quarter 3, which includes the last two months of the experimental period (July and August). Future analyses may reveal the reasons for this pattern, but we cannot offer an empirically-based explanation at this time.

12 Unless indicated, differences on the baseline survey between respondents in control and treatment stores are not statistically significant. The response rate is 49.5% for the pre-intervention survey and 51.3% for the post-intervention survey. Response rates for associates in control and treatment stores are not significantly different in either wave. In both control and treatment stores, respondents were more likely than non-respondents to be female, white, older than 20, and with more than 6 months seniority. In the 8 weeks prior to fielding the surveys, analyses of data from company systems indicate that respondents and non-respondents did not vary significantly in the consistency of their weekly hours, but non-respondents had more consistent shift timing than respondents. With respect to predictability, however, respondents have significantly higher means than non-respondents across almost all of our measures. We will take these sources of non-response bias into account in future analyses of the survey data.

13 For instance, consistency of shift timing is calculated as the probability that any two randomly selected shifts (s1, s2) have their midpoint in the same block of time (t1=t2). If an employee is always scheduled to start and end at the same time, then their shift timing would be perfectly consistent (Pr(t1=t2) = 1).

14 We distinguish between four time slots: early morning (midpoint 6–10 a.m.), day (midpoint 11 a.m.–3 p.m.), evening (midpoint 4–9 p.m.), overnight (midpoint 10 p.m.–5 a.m.).

15 As a reminder, most treatment stores in San Francisco entered the Intervention during Quarter 4 of 2015 and the roll out in Chicago began in February 2016.

16 Comparing intervention and non-intervention weeks overall, we find that starting times are 14 percent more consistent (0.336 intervention vs. 0.294 non-intervention, $p<.01$) and ending times are 16 percent more consistent (0.352 intervention vs. 0.303 non-intervention, $p<.01$).

17 Note though that end times are significantly more consistent in treatment than in control stores during the baseline period (2015-Q3), suggesting that part of the greater consistency in end times in treatment stores may be due to pre-existing differences.

18 In the post-intervention survey, 67% of associate respondents (or 70% of part-time respondents) from treatment stores agreed or strongly agreed "I generally begin work about the same time each work day," compared to 60% from control stores (or 59% of part-time respondents) ($p<.05$). Similarly, 62% of treatment store respondents (65% of part-timers) agreed or strongly agreed "I generally end work about the same time each work day," compared to only 55% of control store respondents (55% of part-timers) ($p<.1$).

19 Although seasonal variation may help account for differences in work hours pre and post the start of Part-time Plus, selection into this status was staggered over the course of the experiment. Of the 77 associates designated Part-time Plus, most received this status near the start of the Intervention in their store, so November for stores in San Francisco and February for stores in Chicago. Over a third (36%) began in March or April as some stores expanded their Part-time Plus workforce.

20 These numbers exclude the two treatment stores that did not implement this component of the Intervention.

21 Initial multivariate analyses indicate that the lower average hours worked during intervention versus non-intervention weeks is unlikely due to concentrating hours on Part-time Plus associates. There is little evidence that when Part-time Plus hours go up, average hours for other part-time associates go down.

22 These statistics are conservative estimates of coverage because they only reflect coverage through the app. Associates may have picked up posted shifts through other means of communication.

23 Better conversion means that a higher percentage of customers who entered stores bought something. Higher basket values mean that customers spent more per visit.

24 Interviews were not audio recorded, so the quotes in this report are as captured in the notes of the interviewers.

WORKLIFE LAW
UNIVERSITY OF CALIFORNIA
HASTINGS COLLEGE OF THE LAW

THE UNIVERSITY OF CHICAGO
SCHOOL OF SOCIAL SERVICE ADMINISTRATION

UNC KENAN-FLAGLER BUSINESS SCHOOL