View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by School of Hotel Administration, Cornell University

Cornell University School of Hotel Administration

The Scholarly Commons

Center for Hospitality Research Publications

The Center for Hospitality Research (CHR)

4-2-2004

Workforce Scheduling: A Guide for the Hospitality Industry

Gary M. Thompson Ph.D.
Cornell University, gmt1@cornell.edu

Follow this and additional works at: https://scholarship.sha.cornell.edu/chrpubs

Part of the Hospitality Administration and Management Commons

Recommended Citation

Thompson, G. M. (2004). Workforce scheduling: A guide for the hospitality industry [*Electronic article*]. *Cornell Hospitality Report, 4*(6), 7-55.

This Article is brought to you for free and open access by the The Center for Hospitality Research (CHR) at The Scholarly Commons. It has been accepted for inclusion in Center for Hospitality Research Publications by an authorized administrator of The Scholarly Commons. For more information, please contact hotellibrary@cornell.edu.

If you have a disability and are having trouble accessing information on this website or need materials in an alternate format, contact web-accessibility@cornell.edu for assistance.

Workforce Scheduling: A Guide for the Hospitality Industry

Abstract

Creating a workforce schedule that ensures appropriate service levels is a key management function. The many complexities of scheduling can be captured through a process that comprises the following four major steps: (1) forecasting demand, (2) translating the demand forecast into employee requirements, (3) scheduling the employees, and (4) controlling the schedule as the day unfolds. Each of those steps involves its own set of tasks. To create a forecast, a manager must determine what needs to be done to meet the expected demand for a given planning period. While a planning period may be of any duration, a 15-minute period is an effective one to use. In particular, the manager must identify the demand drivers and assess whether they are time variant (that is, variable over short periods) or time invariant (relatively stable over short periods). Another part of the forecasting step is determining the tasks to be done in a given period. Some of the tasks (notably, those involving direct customer service) are uncontrollable, because they must be done on the spot. Other tasks, though, such as side work, are controllable because they can be performed at any time (within reason). Having created a fairly reliable estimate of demand, the manager must next translate that demand into the number of workers needed, using an economics-based labor standard. At this point, the manager is ready to construct a schedule that will do the best job of deploying the staff to achieve the desired economic standards without overstaffing and inflating costs. Scheduling is subject to hard constraints, or factors that must be addressed (such as the number of hours an employee can work in a day), and soft constraints, or factors that are desirable in a schedule but not essential (such as employees' desires for when they work and what tasks they perform). Having created a schedule that will meet the economic standards within the constraints, a manager must finally monitor and fine tune the schedule as the day goes on. Most critically, the manager must decide early on whether the demand estimate for the day is correct—meaning the staffing levels will be sufficient—or whether the actual demand is different from the estimate. If the demand estimate proves incorrect, the manager must further decide whether to take such long-lived actions as calling in workers to take care of a big day (or send them home if business has died off) or merely take a short-lived action (such as sending employees on break) to account for momentary fluctuations in actual demand. Computer applications can assist managers in most of the workforce-scheduling tasks, but a manager needs to understand the process if only to judge whether the application in question is providing solutions that are reasonably close to the optimal schedule.

Keywords

workforce scheduling, labor scheduling, forecasting, labor drivers

Disciplines

Business | Hospitality Administration and Management

Comments

Required Publisher Statement

© Cornell University. This report may not be reproduced or distributed without the express permission of the publisher

This article is available at The Scholarly Commons: https://scholarship.sha.cornell.edu/chrpubs/192

CORNELL

CHR Reports

Workforce Scheduling:

A Guide for the Hospitality Industry

by Gary M. Thompson, Ph.D.

The Center for Hospitality Research
AT CORNELL UNIVERSITY

CHR *Reports:*
Workforce Scheduling:
A Guide for the Hospitality Industry

is produced for the benefit of the hospitality industry by
The Center for Hospitality Research at Cornell University

Gary M. Thompson, Executive Director
Glenn Withiam, Director of Publication Services

Advisory Board

Roger Cline, *Chairman and CEO,* Roundhill Hospitality

Richard Cotter, *EVP, Hotel and F&B Operations,* Wynn Resorts

Bjorn Hanson, Ph.D., *Global Hospitality Industry Managing Partner,* PricewaterhouseCoopers

Unmesh Joshi, *Chairman and Managing Director,* Kohinoor Group

Jo-Anne Kruse, *SVP, Human Resources,* Cendant International

Craig Lambert, *President,* Fieldstone Group

Mark V. Lomanno, *President,* Smith Travel Research

Gordon S. Potter, Ph.D., *Associate Professor,* Cornell University

Janice L. Schnabel, *Senior Vice President,* Marsh's Hospitality Practice

David A. Sherf, *SVP, Real Estate and Investment Analysis,* Hilton Hotels Corporation

Judy A. Siguaw, D.B.A., *J. Thomas Clark Professor of Entrepreneurship and Personal Enterprise,* Cornell University

Barbara Talbott, *EVP Marketing,* Four Seasons Hotels and Resorts

Elaine R. Wedral, *President,* Nestlé R&D Center and Nestlé PTC New Milford

R. Mark Woodworth, *Executive Managing Director,* The Hospitality Research Group

Peter Yesawich, Ph.D., *President and CEO,* Yesawich, Pepperdine, Brown & Russell

Workforce Scheduling
is CHR *Reports*, Vol. 4, No. 6 (April 2004)

Single copy price US$50.00
Copyright © 2004 by Cornell University

Thank you to our generous corporate supporters!

Senior Partner

Nestlé

Nestlé

Partners

AIG Global Real Estate Investment

Bartech automatic systems

CENDANT

Kohinoor

MARSH The world's #1 risk specialist™

AIG Global Real Estate Investment

Bartech Systems International

Cendant Corporation

Cornell Hotel Society Foundation

Four Seasons Hotels and Resorts

Kohinoor Group

Marsh's Hospitality Practice

Sponsors

Willowbend

WYNDHAM INTERNATIONAL

Willowbend Golf Management

Wyndham International

Friends

ARAMARK • DK Shifflet & Associates • ehotelier.com • Global Hospitality Resources, Inc. • Hospitality World • hospitalitynet.org • Hotel Asia Pacific • Hotel China • Hotel Interactive, Inc. • Hotel Resource • International CHRIE • The Lodging Conference • Lodging Magazine • Lodging Hospitality • National Hotel Executive Magazine • Resort+Recreation • RestaurantEdge.com • Shibata Publishing Co. Ltd. • Smith Travel Research • The Hospitality Research Group of PKF Consulting • TravelCLICK • UniFocus • WiredHotelier.com

CORNELL The Center for Hospitality Research

Executive Summary

Workforce Scheduling:

A Guide for the Hospitality Industry

by Gary M. Thompson, Ph.D.

Creating a workforce schedule that ensures appropriate service levels is a key management function. The many complexities of scheduling can be captured through a process that comprises the following four major steps: **(1)** forecasting demand, **(2)** translating the demand forecast into employee requirements, **(3)** scheduling the employees, and **(4)** controlling the schedule as the day unfolds. Each of those steps involves its own set of tasks. To create a forecast, a manager must determine what needs to be done to meet the expected demand for a given planning period. While a planning period may be of any duration, a 15-minute period is an effective one to use. In particular, the manager must identify the demand drivers and assess whether they are time variant (that is, variable over short periods) or time invariant (relatively stable over short periods). Another part of the forecasting step is determining the tasks to be done in a given period. Some of the tasks (notably, those involving direct customer service) are uncontrollable, because they must be done on the spot. Other tasks, though, such as side work, are controllable because they can be performed at any time (within reason). Having created a fairly reliable estimate of demand, the manager must next translate that demand into the number of workers needed, using an economics-based labor standard. At this point, the manager is ready to construct a schedule that will do the best job of deploying

the staff to achieve the desired economic standards without overstaffing and inflating costs. Scheduling is subject to hard constraints, or factors that must be addressed (such as the number of hours an employee can work in a day), and soft constraints, or factors that are desirable in a schedule but not essential (such as employees' desires for when they work and what tasks they perform). Having created a schedule that will meet the economic standards within the constraints, a manager must finally monitor and fine tune the schedule as the day goes on. Most critically, the manager must decide early on whether the demand estimate for the day is correct—meaning the staffing levels will be sufficient—or whether the actual demand is different from the estimate. If the demand estimate proves incorrect, the manager must further decide whether to take such long-lived actions as calling in workers to take care of a big day (or send them home if business has died off) or merely take a short-lived action (such as sending employees on break) to account for momentary fluctuations in actual demand. Computer applications can assist managers in most of the workforce-scheduling tasks, but a manager needs to understand the process if only to judge whether the application in question is providing solutions that are reasonably close to the optimal schedule.

CHR Reports

FOR THE

Hospitality Industry

www.chr.cornell.edu

When the Lights Went Out: Hotel Managers' Perceptions of the Blackout of '03, *by Robert J. Kwortnik, Ph.D.*

Restaurant Revenue Management, *by Sheryl E. Kimes, Ph.D.*

Understanding Switchers and Stayers in the Lodging Industry, *by Iselin Skogland and Judy Siguaw, D.B.A.*

Evolution in Electronic Distribution, *by Bill Carroll, Ph.D., and Judy Siguaw, D.B.A.*

Also of interest:

***Tools for the Hospitality Industry #1:* Turnover Cost Evaluator**, *by Tim Hinkin, Ph.D., and Bruce Tracey, Ph.D.*

***Tools for the Hospitality Industry #2:* MegaTips: Scientifically Tested Ways to Increase Your Tips**, *by Michael Lynn, Ph.D.*

CORNELL

The Center for Hospitality Research

AT CORNELL UNIVERSITY

Workforce Scheduling: A Guide for the Hospitality Industry

by Gary M. Thompson, Ph.D.

In the late 1990s the *Cornell and Hotel Restaurant Administration Quarterly* (*CQ*) published a four-article series that I wrote on workforce scheduling.[1] This CHR Report is a synthesis of those articles and also a high-level primer for workforce scheduling.

Workforce scheduling is at once the most essential and the most complex task facing a hospitality manager. The goal of workforce scheduling is to match the number of workers available with the customer demand that exists in any given time period. Thus, a manager must forecast demand, translate forecasts of customer demand into desired staffing levels, and then schedule the appropriate number of employees (neither more nor less than needed) to meet the forecasted demand. Moreover, the schedule must make a reasonable effort to accede to employees' preferences for work times and days. Once operations have started, the manager must monitor the effectiveness of the forecast and the resultant schedule to see that demand is, in fact, being met. If the manager notes excess employees for the demand, some workers can be reassigned, put on break, or sent home. If the demand exceeds the existing staff's capacity, the manager may need to call in more workers on the spot.

Thus it is that cost-effective workforce scheduling is one of the most important tasks that front-line managers perform in service organizations. Hospitality businesses are typical of service operations in that their labor costs constitute a large portion of the costs under managers' control. Controlling costs through effective workforce scheduling can be a challenge, particu-

[1] See: G.M. Thompson, "Labor Scheduling, Part 1: Forecasting Demand," *Cornell Hotel and Restaurant Administration Quarterly*, Vol. 39, No. 5 (October 1998), pp. 22–31; G.M. Thompson, "Labor Scheduling, Part 2: Knowing How Many On-duty Employees to Schedule," *Cornell Hotel and Restaurant Administration Quarterly*, Vol. 39, No. 6 (December 1998), pp. 26–37; G.M. Thompson, "Labor Scheduling, Part 3: Developing a Workforce Schedule," *Cornell Hotel and Restaurant Administration Quarterly*, Vol. 40, No. 1 (February 1999), pp. 86–96; and G.M. Thompson, "Labor Scheduling, Part 4: Controlling Workforce Schedules in Real Time," *Cornell Hotel and Restaurant Administration Quarterly*, Vol. 40, No. 3 (June 1999), pp. 85–96.

Four Steps of Workforce Scheduling

This report will explain the four steps of workforce scheduling. Those steps are:

(1) Forecast customer demand. Predict customer demand for your service by forecasting characteristics of the service transaction that change over time, such as customer-arrival rates.

(2) Translate those demand forecasts into employee requirements. Calculate the number of employee hours required to satisfy demand predicted in step one. This requires setting the number of employees with appropriate skill levels that are needed to serve customers adequately during the time period in question.

(3) Schedule employees. Develop the actual work schedule by taking into account employees' skills, desires, and requests, and then deciding who will do what work at what time.

(4) Fine tune the schedule in real time. Change the work arrangement as required by actual demand. This final step ensures effective customer service.

larly since no two employees have exactly the same skills or desire to work the same number of hours, and because the manager must also heed government regulations, company policies, and contractual obligations. At any given moment, having too few employees—or enough employees but not those who have the necessary skills—can result in poor customer service; frustrated, overworked employees; and lost sales. On the other hand, having too many employees either **(1)** reduces operating margins, if extra hours are scheduled or **(2)** results in low employee morale if employees work fewer hours than they desire because the available work is spread thinly among many employees.

Structure of the report. This report is divided into five parts. In this, the first section, I define workforce scheduling, tell you why I think it's important, define terms, synthesize my *CQ* articles, describe the characteristics I see as important to look for when selecting a workforce scheduling system, talk about what I see as the potential pitfalls and major opportunities of workforce scheduling, and make some personal observations about trends in workforce scheduling. Sections two through five of the report cover the four component tasks in detail.

Intended audience. The intended audience for this report includes senior executives who want an overview of the issues related to workforce scheduling. Such readers will find the first section of the report to be the most useful and will probably want to skip the other sections. Another intended audience is technically competent managers responsible for workforce scheduling in their organizations. These readers, in particular, should find sections two through five quite helpful.

Basis in research and management practice. This report distills the lessons learned from 15 years of research on workforce scheduling (which resulted in approximately 19 published papers on the topic) combined with my experience with actual scheduling problems. Prior to writing the four-paper series for the *CQ*, I had developed a "black box" scheduler for a prominent hospitality company that still uses it each week for scheduling tens of thousands of employees.[2] Since writing the series, I have served as a consultant to a start-up company that offered a web-based workforce-scheduling solution. For the past six years I have been developing

[2] This scheduler was integrated with a front-end graphical user interface provided by another vendor.

specialized workforce-scheduling software for educational institutions (and selling that software commercially for five years). Collectively these experiences have given me a deep appreciation for the true challenges of workforce scheduling.

Section 1

What Is Workforce Scheduling?

Simply put, workforce scheduling, which is also known as labor scheduling, involves putting the right people on the right jobs at the right times (with regard to customer demand). As I stated in the first *CQ* article, workforce scheduling has the following four component tasks, which are highlighted in the box at left. They are:

(1) forecasting customer demand, **(2)** translating those demand forecasts into employee requirements, **(3)** scheduling employees, and **(4)** fine-tuning the schedule in real time.

The first task is to predict customer demand for your service. This initial step involves forecasting characteristics of the service transaction that change over time, such as customer-arrival rates. The second task is to calculate the number of employee hours required to satisfy the demand predicted in step one. In other words, step two requires setting the number and skill levels of employees needed to serve customers adequately during a particular time period. The third task is to develop the actual work schedule by taking into account employees' skills, desires, and requests, and then deciding who will do what work at what time. The final task involves changing the work arrangement as required by actual demand. This final step ensures effective customer service.[3]

[3] Thompson (Part 1), p. 23.

Why Managers Should Care about Workforce Scheduling

As I see it, managers have three primary reasons to care about workforce scheduling, starting with employee preferences. Employees generally have distinct preferences regarding their job, including the tasks to which they are assigned, when and how long their breaks are, with whom they take breaks, with whom they are working, the time of day they work, which days off they have, and whether their days off are consecutive.

Workforce scheduling involves putting the right people on the right jobs at the right times to meet customer demand.

These preferences are commonly complementary across employees. By "complementary preferences" I mean that employees' different work preferences and characteristics tend to offset each other. Employees differ in the characteristics they want in a schedule, and they differ in the importance they place on those characteristics. For example, one employee may hate working on weekends, while another might love weekend work and wants some weekdays free. Although the preferences are not 100-percent complementary—in other words, there may not be equal numbers of people who want the opposite type of schedule—just knowing who wants what and using that knowledge will give you a work schedule that comes reasonably close to meeting employees' desires, which presumably translates to better on-the-job performance and better customer service.

The second reason to care about workforce scheduling relates to the actual time spent on developing a labor schedule. Some firms use what might be called the "photocopier" method of developing a labor schedule, which entails creating this week's schedule simply by duplicating last week's schedule. Although this is a time-efficient approach, the schedules it yields rarely have any practical value, since the photocopier method does not adapt to changes in customer demand or to changes in employee availability or preferences. Another method of developing the schedule is for a manager to build it on a weekly basis, drawing on his or her experience about what will be necessary to serve customer demand and his or her knowledge about the preferences and availability of his or her employees. Though this method may actually yield good schedules, it can require a substantial time commitment for a conscientious and thoughtful manager. Unfortunately, the time a manager spends developing a schedule leaves him or her less time for actually managing the employees and interacting with customers. Yet another approach would be to automate the schedule-development process, though this requires a computer system that can both forecast demand and account for all the employee idiosyncrasies with which adept managers have to deal. Computer systems of this kind are the holy grail of workforce scheduling.

Profitability and effectiveness together are the third reason to care about workforce scheduling. Because a good labor schedule has the right people working the right jobs at the right times, good labor schedules deploy labor in the most effective manner. Effectively deploying labor translates to higher profitability, because short-term overstaffing and long-term understaffing are reduced. The resulting better, more-consistent customer service translates to more future business. A related benefit is that good workforce scheduling allows upper management to monitor performance more closely, both within and across units. Consistency in labor deployment across units is particularly important in chains, because inconsistent labor deployment is one of the main reasons why service quality varies across units.

Terms

A *scheduling horizon* is the period for which schedules are developed at one time, typically one week to several months. *Planning periods*, which are also called planning intervals, are subsets of the scheduling horizon and are the detailed intervals used for staff planning. Planning interval durations of 15 to 30 minutes are commonly used in service industries. *Overstaffing*, or surplus staffing, is a situation where more staff is scheduled in a planning period than is ideally needed. *Understaffing*, or short staffing, is a situation where fewer staff is scheduled in a planning period than is ideally needed.

Major Component Tasks

Exhibit 1 details the major component tasks in workforce scheduling. It also categorizes the tasks according to the primary task of which the component is a part and the frequency with which the component task should be repeated.

Selecting a Workforce-scheduling System[4]

Hospitality firms have many choices for workforce-scheduling systems. Windows-based workforce-scheduling software can be purchased for as little as several hundred dollars, or firms can

[4] Adapted from: *Ibid.*

Exhibit 1

Significant component tasks in workforce scheduling

Primary task	Sub tasks	Frequency
Forecast customer demand (Forecast)	Determine the nature of the work's timing flexibility (controllable or uncontrollable). Controllable work is work for which there is some degree of timing control over when it can be performed. It typically can proceed without having customers present. Examples of controllable work are prep work in restaurants and housekeeping in hotels. Uncontrollable work is work over which there is little if any timing control, since it must be done when customers are in the system. Examples include serving customers in restaurants and checking in guests in hotels.	Infrequently
	Identify those factors that generate the work. These factors are the key labor drivers. An example of a key labor driver for restaurant waitstaff would be the number of parties to be served.	Infrequently
	Determine whether the key labor drivers are time variant, meaning they vary over short time periods, or time invariant, meaning they are relatively stable over short time periods.	Infrequently
	Determine the time interval for tracking the time-variant labor drivers. This task identifies whether you should track the labor drivers using periods of, say, 60 minutes, 30 minutes, 15 minutes, or periods of some other duration. In general, time intervals of 15 minutes or shorter tend to work best.[1]	Infrequently
	Develop forecasts of the time-variant labor drivers, using historical data. These forecasts can be modified, if desired, based on managers' special knowledge of future events.	Weekly
	Reduce the random variation in the forecasts of the time-variant labor drivers by smoothing the forecasts across future periods. This step is important because the goal is to staff to the true level of customer demand and not staff to random variation in demand.	Weekly
	Measure the accuracy of the forecasts of the time-variant labor drivers. Forecast accuracy affects the number of employees needed to serve customers. Inaccurate forecasts mean that you need more staff, which would be considered excess if the forecasts were more accurate.	Weekly
	Determine the time period in which the controllable work can be performed. Some controllable work is itself dependent on the forecasts of time-variant labor drivers. For example, prep work in a restaurant kitchen, which is controllable work, must be done within time limits that are, in part, imposed by when customers will arrive.	Weekly
Translate customer demand into employee requirements (Translate)	Select a labor standard. The choices are productivity standards (which deliver a consistent level of labor use), service standards (which deliver a consistent level of service), and economic standards (which deliver a consistent level of financial performance).[2]	Infrequently
	Determine the number of persons of each skill level needed for each time interval in the week. Do this using the selected labor standard, the forecasts of time-variant labor drivers, and information on employee productivity and the accuracy of the forecasts.	Weekly
Build the schedule (Schedule)	Select a scheduling framework. The scheduling framework, of which several are common, is the paradigm used to represent the workforce-scheduling problems.[3]	Infrequently
	Construct a labor schedule. Do this using information on employee availability, skills, and preferences and the information on the number of employees needed in each period (see the accompanying text for a discussion of the characteristics of good workforce scheduling systems).	Weekly
Modify the schedule in real time (Control)	Monitor the schedule in real time, looking for capacity–demand imbalances. A capacity–demand imbalance exists when either more or less labor is available than is needed at a particular moment. The number of employees needed at a particular point would be determined based on the chosen labor standard, the current actual customer demand, and the productivity of the currently available employees.	Real time
	When a capacity–demand imbalance is detected, make a judgment as to whether it is likely to be short lived or long lived.	Real time
	If the capacity–demand imbalance is judged to be short lived, take actions like extending shifts, sending employees on breaks, or recalling employees from breaks.	Real time
	If the capacity–demand imbalance is judged to be long lived, take actions like calling additional employees in to work or sending employees home early.	Real time

Notes: [1] G.M. Thompson, "Planning Interval Duration in Labor Shift Scheduling," *Cornell Hotel and Restaurant Administration Quarterly* (forthcoming). [2]As explained in the accompanying text, the economic standard is the superior approach to determining schedule parameters. [3] See, for example: G. B. Dantzig, "A Comment on Edie's 'Traffic Delays at Toll Booths,'" *Operations Research*, Vol. 2, No. 3 (1954), pp. 339–341; E.G. Keith, "Operator Scheduling," *AIIE Transactions*, Vol. 11, No. 1 (1979), pp. 37–41; G. Thompson, "Labor Scheduling Using NPV Estimates of the Marginal Benefit of Additional Labor Capacity," *Journal of Operations Management*, Vol. 13 (1995), pp. 67–86; and G. Thompson, "Labor Staffing and Scheduling Models for Controlling Service Levels," *Naval Research Logistics*, Vol. 44 (1997), pp. 719–740.

invest in customized systems costing hundreds of thousands of dollars. I believe that good scheduling systems have the following characteristics.

An intuitive graphical interface. The interface should facilitate tinkering with or editing a schedule. Although good systems will generate schedules that require a minimum of adjustment, systems rarely incorporate all relevant factors, and managers usually must make changes.

A good scheduling engine. The scheduling engine—the algorithms that actually develop the schedule—is harder to evaluate than the graphical interface because one cannot examine the engine directly. (I suggest a remedy for this problem below in the section on testing the autoscheduler.) A good scheduling engine will incorporate effective logic. It should use a cross-period paradigm—one that considers the interactions of staffing decisions across planning periods—rather than a single-period paradigm. It should be holistic (meaning that it develops a schedule while considering employee availability) instead of shift-based (meaning that a schedule is first developed without regard to employee availability and then employees are matched to shifts). The engine should operate quickly and provide schedules that need a minimum amount of tinkering.

Employee preferences. A good scheduling system will require that employees identify their work preferences in advance, including ranking those preferences or identifying tradeoffs among their preferences. The best system will exploit complementary preferences when they exist. By explicitly considering each employee's preferences, a good scheduling system will deliver the best possible schedule in terms of matching the number of employees scheduled to the ideal number of employees needed while at the same time satisfying employee preferences as much as possible.

Constraints. Hospitality businesses should be aware of their particular hard and soft constraints. Hard constraints are those, like government regulations, that must be observed. Soft constraints, by contrast, are those, like employee preferences, that should be observed when possible. Knowledge of the operation's constraints will allow managers to evaluate the degree of congruence between a scheduling system's hard and soft constraints and those identified by the hospitality firm. How a scheduling system deals with conflicting constraints can make the difference between an average system and a good one. Once again, the system's schedule should not require much managerial tinkering to get it into a usable form.

Flexibility. Good scheduling systems allow a high degree of scheduling flexibility, since flexibility yields better schedules. Good systems should allow various times for controllable work, multiple shift lengths, and a variety of break times. The system should also allow employees to be cross-scheduled within and across shifts. However, high flexibility greatly increases the complexity of a scheduling scenario and, consequently, may noticeably lengthen the time a system requires to develop a schedule.

Costs and benefits. Finally, when evaluating workforce-scheduling systems, one should not neglect the costs and benefits, both monetary and operational, of various systems. Inexpensive, off-the-shelf systems, for instance, will generally require that you fit your business operations to the system. Typically, this means that these systems will soon be abandoned. In contrast, a slightly more expensive, customized scheduling solution will fit the system to your business. Your firm's resources and potential benefits of a system will determine the best approach.

Potential Pitfalls

As with any type of management tool, there are potential pitfalls with workforce-scheduling systems. The following are several to keep on your radar:

- *Using only aggregate measures of performance*. Relying on aggregate measures is perhaps the biggest mistake one can make related to workforce scheduling. An example of a commonly used aggregate measure of performance is labor dollars as a percentage of sales (that is, labor-cost percentage). What's missing from this performance measure is what's happening on a period-by-period basis within a week. If my target is for labor to be 20 percent of sales, I could hit that target by having labor be 30 percent of sales in half of my periods and 10 percent of sales in the other half of the periods (assuming sales were similar across periods). Thus, half the time I would have 50-percent more staff on hand than I needed, and the other half of the time I would have 50-percent fewer staff members than needed. A much better measure of performance would be the percentage of weekly periods in which labor is between 18 and 22 percent of sales.
- *Insufficiently testing the effectiveness of an autoscheduler*. If you're going to use a workforce-scheduling system, then you'll want it to deliver the best possible schedule. Because users do not interact directly with the autoscheduler (the "black box" part of the system that actually develops the schedule), it can be difficult to evaluate the performance—a problem that I mentioned above. The best way to evaluate a part of the system that you cannot see is to create some logical test scenarios. The scenarios can be created in a way where it is obvious (to a perceptive manager) what the best schedule should be. These scenarios can then be fed into the scheduling system to see whether it can identify the best possible schedule. If a system cannot find the best schedule even for a simple scenario containing only a few employees, it is unlikely to perform well when it has to deal with real situations containing perhaps hundreds or thousands of employees. Simply put, if the autoscheduler is not up to snuff, then it's likely that the system won't be used, because the schedules won't be good enough. (You might think that the need for a test of this kind is obvious, but I have found that not one of the over 40 universities using my educational-labor scheduling software has done any testing of this sort.)
- *Not using the system*. If you pay for a good system and it meets your needs, then use it. This may entail

Elements of a Good Scheduling System

- ***An intuitive graphical interface.***
- ***A good scheduling engine.***
- ***Inventory of employee preferences.***
- ***Appropriate treatment of constraints.***
- ***High degree of flexibility.***

spending additional money for training. Investing in training may initially seem hard to justify if you have high turnover, but you can certainly assign the scheduling task to job positions that have less turnover. You might also take actions to reduce turnover.

On the other hand, if you've paid for a good system and done the training, but your system is still not being used, perhaps it isn't as good as you thought. In that case, listen to the users of the system in your organization and then pass that information along to the system vendor. If your current vendor doesn't listen, find one who will. From my own experience, the suggestions of users have been invaluable in improving every aspect of my educational-workforce scheduling software.

- *Overblown claims of system vendors*. This pitfall is particularly insidious. Take a look at the website or promotional materials of any vendor of a workforce-scheduling system and you're likely to see inaccurate claims, particularly regarding their autoschedulers (or optimization engines). In a quick review of some vendors' websites, for example, I found one that claims their system will "optimize the end-to-end workforce-scheduling process" and another that claims to offer "optimized schedules." Optimal, for those who don't know, means "the best possible." In reality, given the state of today's computing power and the complexity of workforce-scheduling problems, the systems typically are not able to find "the best" schedules. However, good systems will find good schedules and may, on rare occasions, actually stumble upon the best schedule.

Opportunities Afforded by Workforce-scheduling Systems

In addition to the three items I discussed at the outset regarding why you should care about workforce scheduling, I see two other opportunities. To me, one of the key opportunities is in cross-utilizing employees in different functions at different times. This is difficult to do if schedules are built manually, when the person constructing the schedule may be familiar only with the areas under his or her direct control. However, workforce-scheduling systems can see the "big picture" and so deploy people in different jobs at different times. This capability, which exists today in workforce-scheduling systems, means less understaffing and less overstaffing—in other words more effective labor deployment.

A second key opportunity involves enhancing the way that employee preferences are treated in workforce-scheduling systems. All of the commercial workforce-scheduling systems I've come across and all of the academic literature on workforce scheduling have dealt with employee preferences from a hierarchical perspective. What this means is that employees develop a hierarchy of what is important to them. However, people often make subtle distinctions among alternatives. The marketing literature on discrete-choice modeling offers a better approach to examining employees' preferences.[5] Discrete-choice modeling involves presenting consumers with two or more bundles of products and services that vary on different attributes and asking respondents to select their preferred choice. Along the same line, a

[5] See: Rohit Verma and Gerhardt Plaschka, "The Art and Science of Customer Choice Modeling: Reflections, Advances, and Managerial Implications," *Cornell Hotel and Restaurant Administration Quarterly*, Vol 44, No. 5–6 (October–December 2003), pp. 156–165.

workforce-scheduling system that presents employees with choices between different possible schedules, and then uses that knowledge about their choices when developing a schedule, would take the employee side of workforce scheduling to the next level of sophistication and performance.

Emerging Trends

The following are trends that seem apparent in the workforce-scheduling domain.

- *Availability of on-line workforce scheduling systems*. As with a number of business functions, the internet has changed the workforce-scheduling landscape. Several vendors now offer web-enabled workforce-scheduling solutions. However, web enabling should supplement the key features of workforce-scheduling systems that I identified earlier and not replace those features. I make this observation because this has not always been the case.
- *More integrated management solutions*. Vendors of workforce-scheduling systems appear to be taking a more holistic approach to the task. This translates into systems' tackling all four of the major tasks related to scheduling and, for example, integrating the workforce-scheduling system into time and attendance systems.
- *Labor outsourcing*. Another area that seems to be receiving more attention is labor outsourcing. If a hotel outsourced its housekeeping functions, for example, it would not have to worry about scheduling those housekeepers. In considering outsourcing, one should apply standard processes for evaluating decisions. In addition, I believe it can be helpful to think of employees as value providers and to ask how they can provide enhanced value, rather than think of the staff as a cost to be minimized. If you are confident that you are deploying your labor as effectively as is possible, only then are you ready to evaluate outsourcing.

Section 2

Workforce Scheduling Step by Step

As explained above, forecasting is the first task in workforce scheduling.

Forecasting Demand (Forecast)[6]

This section of the report focuses on forecasting demand for services. My approach comprises the eight FORECAST steps that are listed in Exhibit 1. Again, those are: **(1)** determine the nature of the work; **(2)** identify those factors that generate the work (i.e., the labor drivers); **(3)** determine whether the key labor drivers vary over a short time period (i.e., are time-variant or time-invariant); **(4)** determine the time interval for tracking the time-variant labor drivers (e.g., 15 minutes, an hour, or an eight-hour work shift); **(5)** forecast the time-variant labor drivers; **(6)** reduce the period-to-period variability of the time-variant labor drivers by smoothing; **(7)** check the accuracy of any forecasts through careful measurement and tracking; and **(8)** define the

If you've paid for a good workforce scheduling system, but it isn't being used, perhaps it's not as good as you think it is.

[6] Adapted from Thompson (Part 1), *op. cit*.

time period during which the work can be actually performed (i.e., the work window). Those eight steps are described in detail below.

Step 1—Determine the Nature of the Work

Hotel managers' workforce-scheduling task involves taking into account that work which can be performed at almost any time (e.g., vacuuming public-area carpets) and work that must be performed on demand (e.g., filling a guest's room-service order). As I mentioned above, I coined the terms "controllable work" to describe those tasks that have substantial time flexibility, and "uncontrollable work" to describe those tasks that have little or no timing flexibility (for example, when customers and employees interact).[7] Much of a hospitality firm's work involves delivering service, which generally involves uncontrollable tasks that can be done only when customers are present. In a full-service restaurant, for example, uncontrollable work would include the interaction of the wait staff with customers and the preparation of the customers' orders by the kitchen staff.

One step in forecasting demand is to identify the labor drivers—the factors that determine the amount of work to be done—and to make sure that they are independent of each other.

In contrast to uncontrollable work, controllable work affords some latitude (called a window) in when it can be performed—guest-room preparation, for instance, usually has a window. The timing of that task is flexible because it can be done anytime between when the room is vacated (window opens) until just prior to its being reoccupied (window closes). In a full-service restaurant, controllable work includes washing dishes, stocking shelves, and folding napkins.

Controllable-work windows vary in length depending on the task. For example, preparing a stay-over guest's hotel room might offer just an eight-hour window, say, from 8:00 AM through 4:00 PM. In contrast, a room that turns over and is not immediately filled may have a much longer preparation window, perhaps stretching across several days depending on the next guest's arrival.

If a job has any controllable characteristics, I treat the entire task as controllable. Some tasks that usually are controllable can become uncontrollable in some circumstances, particularly when full customer capacity is reached. Such uncontrollable work situations constitute a special case of controllable work that has a performance window of zero length. Consider what happens to housekeeping in a hotel having no defined check-in and check-out times as room occupancy measured on a per-minute basis approaches 100 percent. In that case, housekeepers have no latitude regarding when a guest room can be prepared. They must prepare each room immediately upon each guest's departure so that the room can be offered immediately to the longest waiting or next arriving guest.

Managers' scheduling task involves forecasting both uncontrollable and controllable work. For uncontrollable work, managers need to predict the volume of work likely to be generated by

[7] G. M. Thompson, "Improving the Utilization of Front-Line Service Delivery System Personnel," *Decision Sciences,* Vol. 23 (September–October 1992), pp. 1072–1098.

the various labor drivers. Such predictions or forecasts should be expressed in terms of whatever workforce-scheduling intervals, also called planning intervals or planning periods, the organization uses (e.g., 15- or 30-minute work periods) for some specified time into the future, commonly called the scheduling horizon (typically, a week or month). Managers also need to forecast the level of controllable work that must be performed during the same planning periods. Because the timing of that type of work is by definition flexible, however, they need not define specifically when it must be performed. Instead, managers must only identify the earliest and latest times when the work can be performed.

The next six steps are common to both controllable and uncontrollable work. The last step, defining the allowable work window, applies only to controllable work.

Step 2—Identify the Independent Labor Drivers

The second step in forecasting demand is to identify those variables that will affect the number and skills of employees needed to perform or deliver the service. Consider housekeeping in a hotel, for example. What affects the length of time a housekeeper requires to clean and prepare a room? By answering that question, one has identified the labor drivers for room attendants.

Work-measurement studies may help a manager to determine how service characteristics affect the duration and nature of the service transaction. To conduct work-measurement studies, managers and employees (and oftentimes consultants) first brainstorm the characteristics of a job that could affect its duration. Next, one measures and itemizes service interactions' actual duration and characteristics. Finally, the investigator determines the precise effect of each characteristic on the service transaction. Multiple regression can be helpful in this regard.

When identifying labor drivers, it will simplify things to choose measures that are independent of each other. In a quick-service restaurant, for example, both the number of items per order and the order value (i.e., check size) can be thought of as labor drivers. However, the order value will likely be related, or correlated, with the number of items per order. Thus, it would be better to select one of those drivers, but not both. Since one must identify and then forecast the effect of each labor driver, it is preferable to be parsimonious when selecting relevant drivers. One can use a correlation matrix to help identify relationships between labor drivers. (A correlation matrix lists the correlations between all pairs of potential drivers. Some spreadsheet applications have a built-in function to generate correlation matrices.) Any labor drivers with correlations with correlations between -0.5 and 0.5 can be assumed to be independent.

Having identified the set of independent labor drivers, the next step is to determine whether the drivers are time-variant or time-invariant.

Step 3—Determine Whether the Labor Drivers Are Time-Variant or Time-Invariant

Examine each labor driver to determine whether its effects will vary over the course of the planning horizon. The effects of a time-variant labor driver will change over the duration of the planning horizon, while the effects of a time-invariant labor driver will remain constant. In the long run every driver is time variant (because, for instance, technological improvements may

Exhibit 2

Two labor drivers plotted by daily period

Plotting two labor drivers over time shows the extent to which their means vary. With a mean that appears relatively stable over time, Driver B is most likely time-invariant. On the other hand, Driver A, with a mean that appears to vary over time, is probably time-variant—and must enter into a manager's forecast of work needed.

introduce new efficiencies), but our focus is on just the scheduling horizon, typically somewhere between one week and three months. With a short work-schedule time horizon, many labor drivers will be essentially constant.

Identifying the time-variant labor drivers is important because of how one must forecast those labor drivers. Perhaps the easiest way to distinguish between time-variant and time-invariant labor drivers is to track the driver over time.[8] This can be visualized by drawing a graph and charting the driver's effect on the y-axis and time on the x-axis. Having graphed a measure, a time-variant driver's chart will show cyclical changes over time, while a time-invariant driver will remain relatively constant or exhibit random variation.

As an example, again consider housekeeping. If one plots the cleaning time per stay-over room, one should see a relatively constant relationship over a period of several weeks, suggesting a time-invariant labor driver (i.e., each stay-over room requires *x* minutes to clean). A consistent time to clean each room does not imply that the requirement for housekeepers also is constant, however. The key time-variant driver for housekeepers is the *number* of rooms to be processed across the scheduling horizon. Note, then, that different labor drivers for a particular job can be either time-variant (e.g., number of rooms) or time-invariant (e.g., effort per room).

Exhibit 2 shows a plot of two labor drivers by daily planning period. Both drivers exhibit random variation from period to period. Driver B appears to have a stable mean over the day, while Driver A's mean appears to vary over time. Based on the picture recorded in Exhibit 2, one would likely decide that Driver A is time-variant while Driver B is time-invariant.

A correlation analysis can be done to help distinguish between time-variant and time-invariant drivers. The correlation should be measured with the data lagged by one period (that is, one compares each period with the period before it). If the data are time variant, one would expect to see a correlation of approximately 0.5 or higher. By the same token, data with low correlations indicate a time-invariant driver. Lagged-data correlation analyses of the labor drivers illustrated in Exhibit 2 yield a correlation of 0.927 for Driver A and 0.101 for

[8] Ideally managers will have a record of each labor driver's effects so that those effects can be plotted as a direct function of time. If not, the drivers can be treated as constants at first (i.e., as time invariant), although data should be collected to determine the drivers' actual effect as either time variant or time invariant.

Driver B. Those results suggest that Labor Driver A is time variant, while Driver B is time invariant. Again, the distinction between time-invariant and time-variant drivers is important because we want to be able to forecast the time-variant drivers, as described in the next subsection.

Step 4—Determine the Time Interval for Tracking Time-Variant Labor Drivers

Once the time-variant labor drivers have been identified, an appropriate time interval for tracking those labor drivers can be determined.

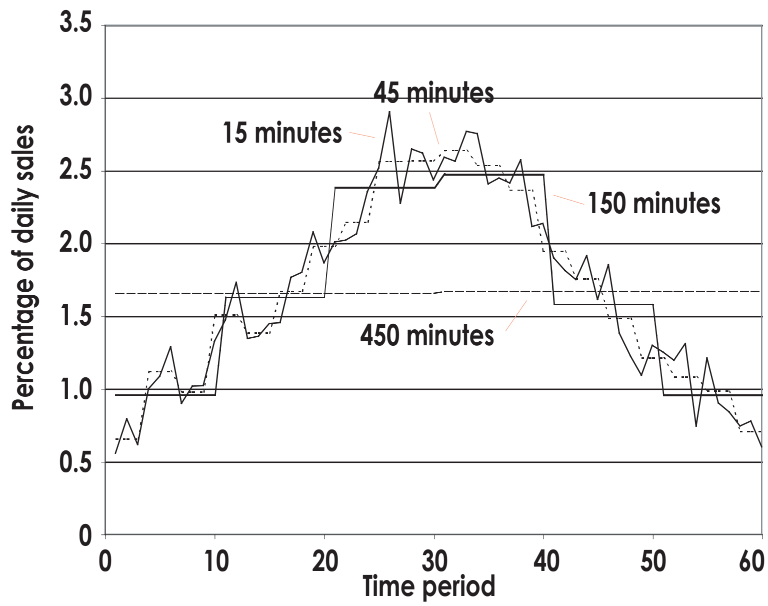
The best way I know to illustrate the process of selecting an appropriate time interval is to use a hypothetical example. Exhibit 3 shows four sets of data for a single labor driver, plotted at four different time intervals, ranging from 15 minutes to 450 minutes. Looking at Exhibit 3, the longest data-collection interval of 450 minutes (7.5 hours) clearly provides a misleading view of the labor driver. In fact, at 450-minute intervals, the flatness of the plotted line suggests that the labor driver is time-invariant. Even the 150-minute and 45-minute intervals in this example also are both too long, because when we compare those data to the 15-minute-interval data it's clear that there are periods where the 150-minute and 45-minute intervals over- and underestimate the level of the driver (essentially smoothing out its actual variability). In this case, then, using those intervals to forecast demand would not give the best results.

In a recent study I found that 15-minute intervals work well for tracking time-variant labor drivers, though periods as short as 5 minutes can be effective when labor drivers are undergoing rapid change (for example, when considering the period between 11:00 AM and 1:00 PM in a quick-service restaurant).[9] Moreover, 15-minute intervals are convenient. They are commonly used anyway when developing work schedules, since employees' rest breaks often are 15 minutes long. On the other hand, tracking time-variant labor drivers using different time intervals is acceptable.

Once the four steps outlined above have been completed, they can be updated only periodically, say, every six months to a year. (One does not need to re-evaluate the chosen time intervals every week, for example.) In contrast,

Exhibit 3

A labor driver plotted according to different data-collection intervals

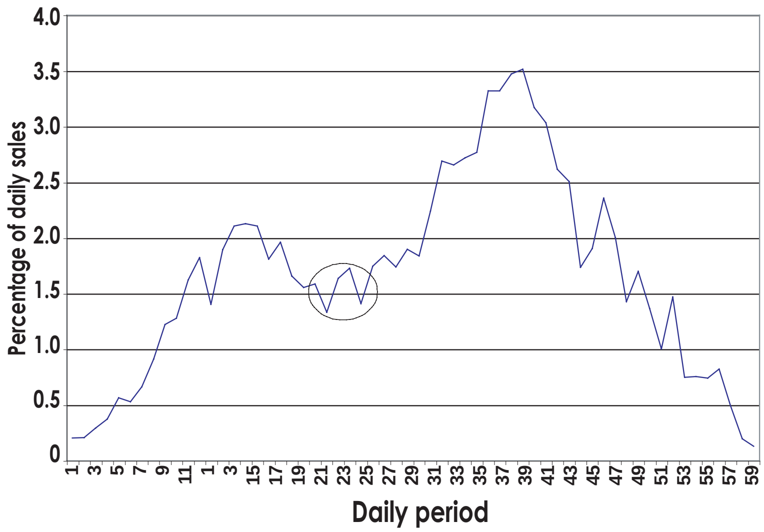


The length of a given data-collection interval influences how much of a given driver's variability is captured. In the example here, the too-long 450-minute interval captures no variability. Yet, as demonstrated by the 15-minute measurements, this driver shows considerable variability.

[9] See: Thompson (2004), *forthcoming*.

Exhibit 4

Forecast of demand, by 15-minute periods



The circled area shows the demand forecast's unreliable "teeth."

the following four steps should be performed every time a new schedule is developed.

Step 5—Forecast the Work Generated by the Time-Variant Labor Drivers

Step 5 involves forecasting the work to be generated by the time-variant labor drivers. To do this, one needs to forecast the level of each time-variant labor driver for every time interval (e.g., every 15 minutes) for the entire work schedule (e.g., for a week or month). There are two ways to make such forecasts: the labor driver can be forecast in each period independently, or the labor driver can be forecast using an aggregation–disaggregation approach. Again, allow me to explain by example.

Assume that a manager wishes to develop a workforce schedule in 15-minute intervals for the coming week and that the facility operates 12 hours per day. Also assume that the key labor driver is the number of customers served and that the workload per customer is time-invariant. The manager, then, must develop forecasts of demand for 336 planning periods (7 days x 12 hours per day x 4 periods per hour = 336). A manager using the independent approach would develop the forecasts independently for each of the 336 periods. A forecast for period *x* would be based on the customer demand observed in the same period during some previous time period, say, six months. One can see that the independent approach requires considerable calculation to generate a large number of forecasts. It also assumes that each period is independent of every other period, which is not true in most hospitality operations. If a hospitality firm is busier than average between 10:00 and 10:15 on a Monday morning, say, it is probably also busier than average between 10:15 and 10:30. To remedy the disadvantages of the independent approach I suggest the aggregation–disaggregation approach to scheduling.

The aggregation–disaggregation approach takes advantage of consistency in the labor driver, a consistency that typically exists in hospitality organizations. It first combines demand data across all the planning periods (in this instance, 15-minute intervals). For example, the manager can combine all the historical Monday demand data and measure how each 15-minute planning period on a given Monday compares to Monday's total business. He can then forecast demand for Monday as a whole (one forecast for each day of the week) and then separate the total day's demand into the demand for each individual period, again using historical data.

One can apply the aggregation–disaggregation approach to longer planning periods as well. For example, the manager can combine the daily demands to obtain a measure that can be used to forecast the weekly demand (one forecast). He would then break down the weekly forecast into the demand for specific days, which in turn would be broken down into demand forecasts for the individual 15-minute planning periods.

For an aggregation–disaggregation approach to work there must be a degree of consistency in the data. For example, Monday must consistently be the *X*th busiest day of the week. Managers can also look for consistency of within-day customer demand over time. Charts and correlation analyses are helpful in evaluating whether the data are, in fact, consistent.

An advantage of the aggregation–disaggregation approach is that a manager, based on his or her knowledge of future events, can inflate or deflate the forecast at the weekly (or daily) level, and then that adjustment will be applied for all periods. Despite this advantage, I recommend that managers compare both the independent and aggregation–disaggregation approaches to forecasting and pick the method yielding the most accurate forecasts (i.e., the lowest forecast error).

Be Wary When Tracking Labor Drivers Using Information Systems

When tracking labor drivers, one should focus on when the driver actually occurs. This matter is particularly critical in restaurants that might rely on point-of-sale (POS) systems to record service times. POS systems do not necessarily provide the correct information. For example, in a quick-service restaurant, the POS system records when service starts and finishes for each customer—that is, when the order is received and when it is fulfilled. As such, one could be tempted to use the POS's check opening as the indicator of when labor was necessary. However, usually there is a lag between when customers arrive at the restaurant and when they are served. That lag, of course, is called waiting in line. The longer customers wait for service, the longer the lag. Most critically, that waiting time will not be consistent across the operating day and, as a consequence, the POS will provide misleading information regarding when service is needed. Using the POS data to define the service demand can, in effect, perpetuate poor service. To correct the problem, a manager would need to supplement the POS information with other data—for example, the number of customers in line or customers' queue time during each period of the day. Knowing that will help the manager estimate when customers actually arrive (versus when they are served).

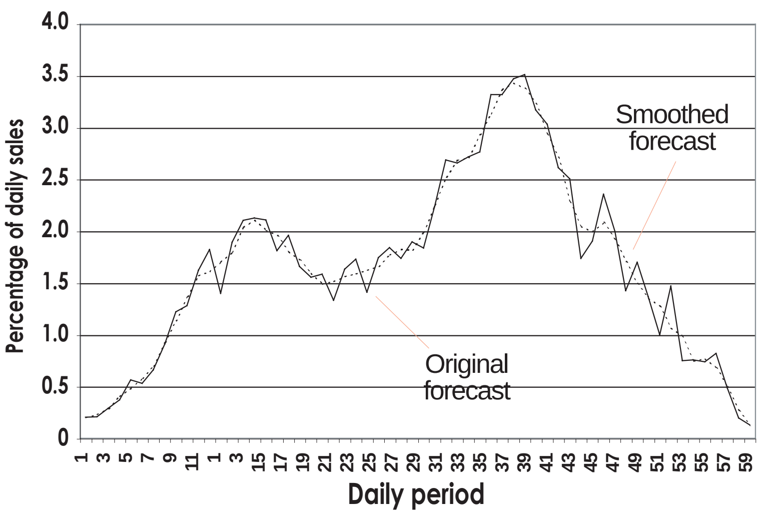
Step 6—Reduce Random Variation by Smoothing

Inevitably, there will be variations in the customer demand. Some of this variation over time is predictable, although clearly some of that change is random and therefore unpredictable. The goal of Step 6 is to deal with both predictable and random change in customer demand over short periods of time. This is best explained by example.

Exhibit 4 illustrates a forecast of sales for a particular weekday (e.g., Mondays), by 15-minute intervals, using four weeks of historical data (in this case, based on the data in Exhibit 3). If that picture (Exhibit 4) represents a good forecast, a manager should be able to explain why she expects demand to materialize this way. In other words, based on the service she delivers to

Exhibit 5

Comparing unsmoothed and smoothed demand forecasts



Smoothing the original forecast (from Exhibit 4 eliminates the uneven "teeth," which are probably caused by erratic data. Caution is needed, though, to ascertain whether the spikes and valleys are caused by real phenomena rather than incomplete data.

customers, in general she should know why customer demand peaks and lags at the times and volumes they do. It is unlikely, however, that she could say exactly what causes the "teeth" in the demand in the circled area of the chart. That is, it's reasonable that she may be unable to explain why the forecast of demand in period 25, for example, is higher than the forecast for period 26, which in turn is lower than the forecast for period 27.

The most likely reason for the "teeth" is that the limited amount of historical data (just four Mondays' worth) results in random variation around the true level of demand at that time of day for that particular day of the week. To eliminate some of that random variation one can smooth the forecast by averaging the original forecast for the period with the original forecasts for the two adjacent periods. Exhibit 5 shows the results of applying the smoothing technique to the data used in Exhibit 4. Note that the smoothed forecast retains the general shape of the original forecast while eliminating the randomness (the teeth) of that forecast.

Smoothing is intended to make up for erratic data. The danger in smoothing forecasts comes when the spikes and valleys of demand are caused by real phenomena, rather than by a lack of data. For example, a hotel's central reservations center may experience a large but short-lived increase in call volume following a television advertisement. If smoothing is applied to the demand trend for that day, the short-duration peak caused by the TV ad will disappear. That, in turn, will result in an understaffed CRS and poor service when the ad campaign runs again.

The determination of whether smoothing is appropriate comes down to understanding the service's various drivers. In the call-center example, managers should know *why* demand peaks occur. In general, if managers cannot offer a valid reason why brief demand peaks exist, either they do not know their service as well as they should or they should control for random variation in their forecasts by using smoothing.

Step 7—Measure and Track Forecast Accuracy

Forecasts are rarely perfect. Step 7 measures and tracks forecast accuracy to ensure that the forecasting method is appropriate. There are two common yardsticks for measuring forecast accuracy: mean absolute percentage error (MAPE) and coefficient of variation of the forecast error (COV). Both

MAPE and COV measure relative error (i.e., actual demand appears in the denominator of both measures). MAPE is found by taking the mean of the absolute value of the error divided by the actual demand, multiplied by 100 percent. COV is found by taking the standard deviation of the error and dividing by the average demand.

In general, the forecast errors should be tracked using the time intervals used for tracking the labor driver. Using the earlier example of a 12-hour-a-day, 7-day-a-week operation, one would measure the forecast error in each of the 336 planning periods. If one plots the forecast errors, one often observes high relative variability at low-demand times. In other words, forecasts commonly are relatively less accurate at low-demand times (albeit, not absolutely so) than they are during high-demand times. This tracking by period is important, since this information is used in the next primary task of translating the forecasts into employee requirements.

Step 8—Define the Allowable Window for the Work

As discussed in Step 1, only controllable work has timing flexibility. Thus, defining a work window applies only to controllable work. A manager must take into account the fact that the window opens at the earliest moment that the job can begin (e.g., a guest vacates a room) and closes at the latest time the job can finish (e.g., moments before the next guest occupies that room). Work windows can be related to the operating hours of the facility or to the controllable work's drivers (e.g., having shelves stocked in a concession prior to the peak sales periods).

Scheduling controllable work requires more information than does uncontrollable work. Instead of just forecasting the labor drivers that define work volume, one also needs to forecast the drivers that define the timing of a particular task. Again consider the housekeeping example. The number of guests arriving on a given day defines the room-preparation workload. The earliest time that the rooms can be prepared is defined by guest departures, while the latest that rooms can be prepared is defined by guest arrivals (or by a hotel's allowed check-in time). So, one needs to forecast both guest arrivals and departures to determine both the volume and timing of the required work.

Drivers that determine controllable work can be either time-variant or time-invariant. For example, the drivers that define the preparation of turned-over rooms are time-variant, depending on the arrival and departure times of guests. In contrast, the driver that defines the allowable time to prepare a stay-over room is time-invariant, since by management policy, those rooms will always be cleaned within certain specified time limits (say, 9:00 AM to 4:00 PM). To forecast the timing window for controllable work, one should follow a procedure much like that outlined in Steps 2 through 7.

The next section of the report describes what one does with the forecasts of customer demand. Specifically, it addresses how the forecasts are translated into desired staffing levels.

Section 3

Translating Forecasts into Staffing Needs (Translate)[10]

This section of the report shows how one might tackle the second primary task, namely, setting the number of employee hours needed to serve customers adequately during a given time period. I refer to this task as

[10] Adapted from Thompson (Part 2), *op. cit*.

TRANSLATE. TRANSLATE uses as its input the forecasts of demand discussed in the previous section. The output of TRANSLATE, the employee requirements, will become the input to the third task—the development of the schedule—which I discuss later.

The output of TRANSLATE depends on whether the work is controllable, as I defined it above.[11] For uncontrollable work, the process of translating demand forecasts into employee requirements should result in the ideal number of employees working in each planning period. That process can also identify the effect of deviating from the ideal staff size. For controllable work the translation process specifies the total workload in labor-hours (or a similar measure) and the window during which the work can occur. As in the example above, the window for turning over guest rooms opens just after a guest departs and closes just before the next guest arrives. The window is based on forecasts of the events that determine those two points.

Three TRANSLATE Approaches

The following are the three basic approaches to translating demand forecasts into employee requirements: using productivity standards, using service standards, and using economic standards.

Productivity standards. The aim of using productivity standards is to define and then rely on consistent (and reasonable) productivity from employees. Productivity standards are the easiest means of translating demand forecasts into employee requirements. An example of a productivity standard in a restaurant might be that a server can handle 14 customers an hour; a productivity standard in a hotel might be that a housekeeper can process 15 rooms a day.

Productivity standards are easily applicable to controllable work. Since an employee performing controllable work does not have to wait for customers to arrive, the employee can work uninterrupted. For example, if housekeepers are scheduled correctly, they can work steadily, since they will not have to wait for guests to check out.

If, however, one cannot exactly predict when customers will arrive (that is, if the work is uncontrollable), one needs to allow for idle time. For example, a server may actually be able to handle 16 customers an hour, yet the productivity standard might be set at one server for every 14 customers. The difference between the maximum service level and the productivity standard is the planned idle time.

The challenge of using productivity standards for uncontrollable work is that planned idle time, and therefore the productivity standard itself, should not be a constant. During times with high customer levels, one can more fully use employees while providing the same level of service. This is possible because having more employees on duty increases the flexibility inherent in the work-management system.

Service standards. The aim of using service standards is to deliver a consistent level of customer service, regardless of the time of day. Service can be measured in many ways. Some possibilities are the average length of time customers wait for service, the average number of customers waiting for service, and the percentage of customers who have to wait more than a specified amount of time for service. Implicit in service standards is the

[11] See: Thompson (1992), *op. cit.*; and Thompson (Part 1), *op. cit.*

recognition that there will be less idle time per employee when the workload is high than when it is low.

The most difficult aspect of developing service standards is determining the appropriate level of service. Customer surveys, focus groups, direct observation, and experimentation can help set the standard.[12]

Economic standards. The aim of using an economic standard is to deliver the appropriate level of service most economically. Typically that means delivering better service at high-demand times than at low-demand times. Better service is economically warranted at high-demand times because it is experienced by more customers. Conversely, the cost of delivering such high service to the few customers arriving in low-demand periods may outweigh the benefits.

Economic standards vary in complexity. A straightforward approach is to estimate the cost of having customers wait for the service. For example, customer-waiting time would be valued much lower in a QSR operation where the average sale is relatively small compared to the value of time for those customers kept waiting on the telephone for reservations agents for an upscale resort hotel. The economic standard would then be used to determine the staffing levels that result in the lowest cost of delivering the service, taking into account labor cost and customer-waiting cost. I discuss a more complex approach later.

Economic standards have been applied in QSRs and telemarketing, with purported success.[13] That success, I believe, arises for two reasons. The more important one is that overall service has been improved. The other is that the service has been tailored to the volume. The major factor—improved service—suggests that service standards are commonly set too low in many businesses that use them.

The ease of implementing economic standards varies. For example, applying economic standards to a hotel's telephone-reservation center would be easier than applying them to its concierge position. One can determine the number of lost calls and estimate the average call value for telephone reservations. It would be difficult, however, to determine the cost of having customers wait for a concierge. At their most complex level, using economic standards requires defining the link between customer-waiting time and long-run (future) business volume. Doing so is the most difficult aspect of using economic standards and may be the reason they are not widely applied.

Comparing the Standards

In one of the *CQ* articles on this topic, I created two scenarios to examine the differences in the three labor standards.[14] These scenarios yielded three important implications. First, though productivity standards or service standards can sometimes match the performance of an economic standard, they do so only across narrow ranges of business volume. That suggests that if the productivity or service standards are to be applied correctly, either they must be applied only in narrow ranges of business volumes or they must change across business volumes. Identifying the relevant business volume ranges and values of the standards would be diffi-

[12] Thompson (1997), *op. cit.*, p. 720.

[13] M. Davis, "How Long Should a Customer Wait for Service?," *Decision Sciences,* Vol. 22, No. 2 (Spring 1991), pp. 421–434; and P. Quinn, B. Andrews, and H. Parsons, "Allocating Telecommunications Resources at L.L. Bean," *Interfaces,* Vol. 21, No. 1 (January 1991), pp. 75–91.

[14] Thompson (Part 2), *op. cit.*

Exhibit 6

Applying an economic standard

Number of Servers	Wq[a]	Total waiting time[b]	Total waiting cost[c]	Labor cost[d]	Total cost[e]
8	2.382	4.447	$44.47	$80.00	$124.47
9	0.722	1.347	$13.47	$90.00	$103.47
10	0.277	0.517	$5.17	$100.00	$105.17

[a] Wq is the average time a customer spends in the queue (from the queuing-model results), in minutes.

[b] Total waiting time is total hours in the queue across all customers, equal to number of customers (112) times the average waiting time, in hours.

[c] Total waiting cost is equal to the total waiting time, in hours, times the estimated hourly cost of customer waiting ($10.00 in this example).

[d] Labor cost is equal to the number of servers times the hourly labor cost per employee.

[e] Total cost is equal to the total waiting cost plus the labor cost.

cult, which is perhaps why their implementation does not result in the anticipated favorable outcome.

Second, productivity standards and service standards are largely inaccurate, even when they are consistent with a particular customer-waiting cost. This inaccuracy leads to staffing levels that are higher or lower than ideal and therefore service-delivery costs that are higher than those resulting from use of the economic standard.

Third, regardless of whether one implements a productivity standard, a service standard, or an economic standard, one is faced with the difficult task of setting an appropriate benchmark for that standard. The problem with productivity standards and service standards is that the labor standards they implement are only surrogates for an economic standard. Since only the economic-standard approach directly addresses the economic effect of good and poor service, it yields better staffing decisions. I therefore contend that hospitality businesses should use economic standards, despite the difficulty of determining the economic value of good and poor service.

Developing an Economic Standard

I've just argued that the economic standard is a better tool than either productivity standards or service standards. To apply this knowledge, however, we must be able to develop that economic standard. Here are some ways to do it. I will illustrate the alternatives using a scenario where employees can serve 16 customers per hour and where 112 customers are expected to arrive in a given hour (necessitating a minimum of eight employees during that hour). Also, I'll assume that the total hourly labor cost (including benefits) is $10.00 per employee, and that the contribution value of a transaction is $5.00 (unless otherwise indicated).

Method 1—*Applying the economic standard.* As noted earlier, the simplest (but not necessarily best) way to implement an economic standard is to estimate a customer-waiting cost. For example, since the transaction contribution is lower than our hourly labor cost per employee, one might initially assume that the per-hour customer waiting cost is approximately equal to an employee's hourly labor cost ($10.00).[15] Using the data shown in Exhibit 6, it is possible to select a staffing level that minimizes the total service-delivery costs. In this case, the ideal number of employees to have on duty for that particular hour is nine.

[15] The chart in Exhibit 6 applies that assumption to the queuing formulas shown in the box on page 28 of Thompson (Part 2), *op. cit.*

Exhibit 7

Applying a revenue focus

Transaction contribution	Number of servers	P(W≤10)[a]	P(W>10)[b]	Transactions made[c]	Transactions lost[d]	Total transaction value[e]	Labor cost[f]	Net benefit[g]
$5.00	8	0.9559	0.0441	107.056	4.944	$535.28	$80.00	$455.28
	9	0.9981	0.0019	111.792	0.208	$558.96	$90.00	$468.96
	10	0.9999	0.0001	111.992	0.008	$559.96	$100.00	$459.96
$100.00	8	0.9559	0.0441	107.056	4.944	$10,705.59	$80.00	$10,625.59
	9	0.9981	0.0019	111.792	0.208	$11,179.18	$90.00	$11,089.18
	10	0.9999	0.0001	111.992	0.008	$11,199.17	$100.00	$11,099.17
	11	1.0000	0.0000	112.000	0.000	$11,200.00	$110.00	$11,090.00

[a] P(W≤10) is the proportion of customers waiting 10 minutes or less (from the queuing-model results).

[b] P(W>10)[b] is the proportion of customers waiting more than 10 minutes (from the queuing-model results).

[c] Transactions made is equal to the number of customers expected that hour (112) times the proportion whose wait doesn't exceed 10 minutes.

[d] Transactions lost is equal to the number of customers expected that hour (112) times the proportion whose wait exceeds 10 minutes.

[e] Total transaction value is equal to the number of transactions made times the transaction contribution.

[f] Labor cost is equal to the number of servers times the hourly labor cost per employee ($10).

[g] Net benefit is equal to the total transaction value minus the labor cost.

One of the problems with the customer-waiting-cost technique is that it assumes a linear relationship between waiting time and cost. For example, it assumes that the cost to the firm of 100 customers each waiting one minute is the same as one customer waiting 100 minutes. The next method offers a way to overcome this shortcoming.

Method 2—*Using a revenue focus.* In switching from a cost to a revenue focus, one attempts to answer the question, how will current and future sales be affected by a wait of x minutes? To apply the simplest form of the revenue-based approach, one need only identify the waiting time at which customers are lost (i.e., after how many minutes' waiting does a customer walk out without making a purchase?), which can be done through observation, experimentation, and experience. For example, let's say that we've observed that making customers wait for less than 10 minutes has no effect on sales, but we lose the current sale for any customer forced to wait more than 10 minutes (loss of future sales is addressed later on). The upper part of Exhibit 7 shows that, in this case, nine servers is the ideal staff for the hour, since the net benefit is maximized at nine servers.

Now, consider the effect of a higher transaction contribution, say, $100. The lower part of the table in Exhibit 7 shows that the ideal number of servers increases to 10. The reason that the tenth server is worthwhile in this example is that the extra server's ability to reduce the number of transactions lost more than offsets the employee's hourly labor cost. In general,

Exhibit 8

Accounting for different customer-waiting times

	Number of servers			
	8	9	10	11
(1) P(W ≤ 3)[a]	0.7146	0.9222	0.9798	0.9951
(2) P(3 < W ≤ 5)[b]	0.1180	0.0510	0.0161	0.0043
(3) P(5 < W ≤ 10)[c]	0.1233	0.0249	0.0040	0.0006
(4) P(W > 10)[d]	0.0441	0.0019	0.0001	0.0000
Transactions lost[e]	15.875	3.022	0.636	0.137
Transactions made[f]	96.125	108.978	111.364	111.863
Total transaction value[g]	$480.63	$544.89	$556.82	$559.32
Labor cost[h]	$80.00	$90.00	$100.00	$110.00
Net benefit[i]	$400.63	$454.89	$456.82	$449.32

[a] P(W ≤ 3) is the proportion of customers waiting three minutes or less (from the queuing-model results).

[b] P(3 < W ≤ 5) is the proportion of customers waiting between three and five minutes (from the queuing-model results).

[c] P(5 < W ≤ 10) is the proportion of customers waiting between five and ten minutes (from the queuing-model results).

[d] P(W > 10) is the proportion of customers waiting more than 10 minutes (from the queuing-model results).

[e] Transactions lost is equal to the number of customers expected (112) times the likelihood of a transaction (customer) being lost; that is, (row 1 x 0) + (row 2 x 0.20) + (row 3 x 0.60) + (row 4 x 1.00) = all lost transactions.

[f] Transactions made is equal to the number of customers expected (112) minus the number of transactions lost.

[g] Total transaction value is equal to the number of transactions made times the transaction contribution (assumed to be $5.00 in this example).

[h] Labor cost is equal to the number of servers times the hourly labor cost per employee ($10).

[i] Net benefit is equal to the total transaction value minus the labor cost.

higher contributions per transaction economically warrant the use of more servers.

The simple revenue-based approach can be further refined. Rather than trying to identify a single cut-off point, beyond which all sales are lost, it is possible to track the proportion of transactions that are lost based on waiting times within certain ranges.[16] For example, let's say that we've observed that for waits up to three minutes no customers are lost, but for waits of three to five minutes we lose 20 percent of the customers, and we lose 60 percent of the customers for waits between five and ten minutes. As before, all sales are lost if the wait exceeds ten minutes. We can then perform an analysis of different staffing levels, as shown in Exhibit 8. The ideal staffing level is that level which returns the greatest benefits—in this case, 10 employees.

The calculation of "transactions lost" in Exhibit 8 merits further elaboration. The transactions-lost figures are found by multiplying the number of expected customers times the likelihood that a customer will leave without making a purchase. The value of the likelihood that a customer will leave is found by summing across the different wait categories the probabilities of customers' having a wait that falls within the category times the observed likelihood that the customer will be lost given that wait. So, in this example, the possible waiting periods are defined as 0 to 3 minutes, 3 to 5 minutes, 5 to 10 minutes, and more than 10 minutes. If there are eight employees on duty, the probabilities for customers' falling into the different waiting periods are 0.715 for 0 to 3 minutes, 0.118 for 3 to 5 minutes, 0.123 for 5 to 10 minutes, and 0.044 for more than 10 minutes. The observed likelihood of losing those waiting customers is 0 for 0 to 3 minutes, 0.2 for 3 to 5 minutes, 0.6 for 5 to 10 minutes, and 1.0 for more than 10 minutes. This translates into an overall likelihood of 0.142 of losing the transac-

[16] Quinn *et al.*, *op. cit.*

Exhibit 9

Accounting for waiting effects on future business

	Number of servers					
	8	9	10	11	12	13
(1) P(W ≤ 0.15)[a]	0.3896	0.6446	0.8033	0.8968	0.9713	0.9759
(2) P(0.15 < W ≤ 3)[b]	0.3250	0.2776	0.1765	0.0983	0.0284	0.0239
(3) P(3 < W ≤ 5)[c]	0.1180	0.0510	0.0161	0.0043	0.0003	0.0002
(4) P(5 < W ≤ 10)[d]	0.1233	0.0249	0.0040	0.0006	0.0000	0.0000
(5) P(10 < W)[e]	0.0441	0.0019	0.0001	0.0000	0.0000	0.0000
Transactions lost[f]	20.819	3.231	0.644	0.137	0.029	0.006
Transactions gained[g]	21.817	36.100	44.987	50.221	53.129	54.651
Net transactions[h]	112.998	144.870	156.343	162.084	165.010	166.645
Total transaction value[i]	$564.99	$724.35	$781.71	$810.42	$825.50	$833.22
Labor cost[j]	$80.00	$90.00	$100.00	$110.00	$120.00	$130.00
Net benefit[k]	$484.99	$634.35	$681.71	$700.42	$705.50	$703.22

[a] P(W ≤ 0.15) is the proportion of customers waiting 0.15 minutes or less (from the queuing-model results).

[b] P(0.15 < W ≤ 3) is the proportion of customers waiting between 0.15 and 3 minutes (from the queuing-model results).

[c] P(3 < W ≤ 5) is the proportion of customers waiting between 3 and 5 minutes (from the queuing-model results).

[d] P(5 < W ≤ 10) is the proportion of customers waiting between 5 and 10 minutes (from the queuing-model results).

[e] P(10 < W) is the proportion of customers waiting more than 10 minutes (from the queuing model results).

[f] Transactions lost is equal to the number of customers expected (112) times the expected number of transactions lost per customer; that is, (row 2 x 0.00) + (row 3 x 0.00) + (row 4 x 0.20) + (row 5 x 0.60) + (row 6 x 2.00) = all lost transactions (due to poor service).

[g] Transactions gained is equal to the number of customers expected (112) times the expected number of additional transactions per customer; that is, (row 1 x 0.50) + (row 2 x 0.00) + (row 3 x 0.00) + (row 4 x 0.00)+ (row 5 x 0.00) = all additional transactions (from good service).

[h] Net transactions is equal to the number of customers expected (112), minus the number of transactions (customers) lost, plus the number of transactions gained.

[i] Total transaction value is equal to net number of transactions times the transaction contribution (assumed to be $5.00 in this example).

[j] Labor cost is equal to the number of servers times the hourly labor cost per employee ($10).

[k] Net benefit is equal to the total transaction value minus the labor cost.

tion, as follows: (0.715 x 0.0) + (0.118 x 0.2) + (0.123 x 0.6) + (0.044 x 1.0) = 0.142. Multiply that value by the 112 customers expected for that hour and the transactions-lost figure equals 15.875 (Exhibit 8).

Davis takes this analysis one step further by attempting to identify the long-term effects of good service and poor service.[17] The argument he uses is that poor service affects not only the current transaction, but it reduces future business. Similarly, good service can serve to increase future business. For example, let's say that, based on

[17] Davis, *op. cit.*

observations, experiments, and experience, we arrive at the following relationship between waiting time and future business:

Wait time in minutes (W)	Effect on future transactions
W ≤ 0.15	0.5
0.15 < W ≤ 3	0.0
3 < W ≤ 5	-0.2
5 < W ≤ 10	-0.6
10 < W	-2.0

The first thing to note from the table above is that serving customers in less than 0.15 minutes (about 10 seconds of wait time) will increase future business by 0.5 transactions, on average, for every customer who experiences that exceptional service standard. The second thing to note is that poor service (a wait over 10 minutes) results not only in the loss of the current sale, but also the loss of another transaction. In other words, poor service results in the loss of two transactions for every customer who experiences that poor service (for example, through the customer's failing to return or through negative word-of-mouth that influences other potential patrons). Once those effects have been estimated, setting the staffing level is straightforward, as shown in Exhibit 9 (on the previous page).

The ideal staff level, then, is that number of employees that can provide the maximum benefit. In this case, 12 employees would be ideal for the hour (as shown by the "net benefit" line of Exhibit 9). In general, the longer the time horizon one considers as being affected by current good or poor service, the higher the appropriate staffing levels. Since estimating future effects can be problematic, managers are likely to underestimate their importance. Again, underestimating the long-run effects can result in a manager's setting staffing levels that are lower than they should be to maximize profits.

Other Issues

Forecast error. When translating demand forecasts into employee requirements, one should take forecast error into consideration. The greater the forecast inaccuracy, the greater the staffing requirements (as shown in Exhibit 10). When there is a high level of forecast inaccuracy, required staffing levels can be as much as 50 percent higher, and service-delivery costs as much as 39 percent higher, than when there are perfectly accurate forecasts. The effects of inaccurate forecasts illustrates the connection between the four tasks of workforce scheduling. Because each task has only a limited ability to correct problems or deficiencies that occurred in an earlier task, a manager must perform each task well.

Work spillover. When translating the demand forecasts into labor requirements one might wonder whether it is safe to do this translation independently for each planning period. In general the answer is no, simply because service spills over from one planning period to another. Consider, for example, a quick-service restaurant that uses 15-minute planning periods, where the mean service duration is three minutes. If the current planning period ends at 1:00 PM, a customer who arrives at 12:50 will be served in the current period. For a customer who arrives between 12:57 and 1:00, however, a portion of the service is performed in each period. The closer a customer arrives to the end of the current period, the greater the demand the customer places for service in the next planning period.

The effect of that spillover is to lag demand. When demand is increasing across periods, the spillover effect reduces the number of staff required in a period, as the service spilling over from preceding periods is less than the service spilling over into the subsequent periods. When demand is decreasing across periods, the spillover effect increases the number of staff beyond what would otherwise be needed. An earlier paper presents a method of calculating the spillover effect.[18]

Deviations from the ideal. Forecasts of demand are translated into employee requirements on a period-by-period basis. When the actual schedule is developed, however, it is often difficult to match the number of employees scheduled to the number of employees needed. There are a number of reasons why this is true, some of which will be covered in the next major section of this report.

If the matching cannot be done exactly, then one would like additional information about deviations from the ideal staff size. That is, in addition to determining the number of employees who should be scheduled, the translation process should also provide information about the cost of deviations from the ideal staff size.

Exhibit 11 provides an example of such information for an economic standard, based on the assumptions of: a labor cost of $10 an hour (including benefits), the ability of employees to serve 16 customers an hour at 100-percent capacity, and the same customer-waiting cost of $13.46 an hour used in scenario 1. In the first period the ideal staff size is five employees. Overstaffing by one employee increases costs $5.27 over the ideal. That is less than the $10 increase in labor costs arising from the additional employee because of the improved customer service and, hence, lower costs for customers waiting.

[18] G.M. Thompson, "Accounting for the Multi-period Impact of Service When Determining Employee Requirements for Labor Scheduling," *Journal of Operations Management,* Vol. 11 (1993), pp. 269–287.

Exhibit 10

Ideal number of employees under increasing levels of forecast inaccuracy

Coefficient of variation of forecast error	Number of employees	Service-delivery cost per hour*
0.00	10	$111.05
0.05	10	$112.81
0.10	11	$117.36
0.15	12	$125.25
0.20	13	$134.46
0.25	14	$144.30
0.30	15	$154.70

*The source of the service-delivery cost in the table above is wages plus costs for customers waiting. Figures are based on the assumptions used in scenario one on pp. 29–30 of Thompson (Part 2), *op. cit.*

Exhibit 11

Cost of deviating from the ideal staff size

Period	Customer-arrival rate	Ideal staff size	-2 employees	-1 employee	+1 employee	+2 employees
1	50.8	5	NA*	$13.72	$5.27	$13.95
2	74.4	7	$122.04	$5.15	$5.69	$14.22
3	118.2	10	$93.62	$8.69	$3.58	$11.00

*It would not be reasonable to have only three employees on duty, because 3 employees x 16 customers per hour = 48 customers per hour maximum (which is less than the minimum customer-arrival rate shown).

If one has to be overstaffed by one employee in one of the periods, it is best to do so in the third period, where the net cost of overstaffing is the lowest. Similarly, if one must be understaffed by one employee in one of the periods, the second period would be the choice. The reason that period three would be the best period to be overstaffed is that in that period the additional person comes closest to covering his or her labor costs.

For controllable work, the economic effect of deviating from the ideal staff size need not be calculated on a period-by-period basis, as controllable work is not attributed to specific periods. Rather, the economic information can be calculated based on deviations over several periods. For example, if one needs a total of 100 hours for housekeeping on a particular day, what is the cost if 99 hours or 101 hours are scheduled? Such information can help one make decisions about how to deploy staff across days and across jobs.

Deploying staff across jobs. Faced with simultaneously staffing two or more jobs, managers often cross-train their employees. The rationale is that it may be possible to use employees more effectively by assigning them to different jobs at different times of the day.

Economic standards offer the only means of directly determining where to allocate staff across jobs. Consider the case where a manager must cover two jobs using a pool of cross-trained employees but has too few employees to staff both jobs at their ideal levels. Neither a productivity standard nor a service standard can offer guidance about which job to short staff. An economic standard, however, allows the manager to make the determination by identifying the economic effect of short staffing for each job.

Feedback. The four tasks of workforce scheduling are closely linked. In particular, the task of translating the demand forecasts into employee requirements is closely linked to the task of developing the workforce schedule and, in fact, should operate with feedback from the scheduling task. A hotel chain's telephone reservation center that I visited can illustrate this point.

In the reservation center, which operates around the clock, the manager used the average productivity of employees in translating demand forecasts into employee requirements. Using average employee productivity in this way is valid if the mix of employees working in all periods is similar. However, that was not the case at this center. The manager developed a generic work schedule and then allowed the employees, in order of their rank, to choose when they wanted to work.

The more senior, more productive employees chose first, and most of them selected day shifts, leaving the less productive employees to work the night shifts. Since the schedule-translation process did not make use of information about the varying productivity of the employees working at different times of the day, staffing levels were not set appropriately. The center could easily have corrected the situation by applying different productivity levels at different times of the operating day.

Absenteeism. A manager should consider absenteeism when developing the employee requirements. It is important to track absenteeism by time of the day, day of the week, and specific job; it does not occur consistently. (In the long run, of course, one should address the causes of systematic absenteeism.) The effect of absenteeism is similar to that of forecast inaccuracy. The higher the absenteeism, the higher the staffing levels should be. Ideal staffing levels should be adjusted upward to reflect the likelihood of absenteeism.

The Best Approach

In summary, comparing the three approaches to translating demand forecasts into employee requirements, an economic standard is the most appropriate for both controllable and uncontrollable work. For uncontrollable work the translation process identifies the ideal number of employees to have working in each of the planning periods in the scheduling horizon. For controllable work the translation process identifies the labor-hour requirement for the work and the window in which the work can be performed. For both kinds of work an economic standard can provide information about the economic consequences of deviating from the ideal staff size.

Additionally, I have described how forecast inaccuracy and absenteeism both increase the number of employees necessary. There is a clear incentive, then, to develop accurate forecasts and to control absenteeism. The outputs of the translation process become inputs to the third step in the process—the development of the labor schedule. The next major section of this report describes that task.

Section 4

Developing the Labor Schedule (Schedule)[19]

This section focuses on the third task, developing the workforce schedule, which I refer to as Schedule. In this section I first present terms and then an overview of workforce scheduling. Next, I identify approaches and techniques for generating workforce schedules and discuss considerations of workforce scheduling. Finally, I present guidelines for selecting a workforce-scheduling system.

[19] Adapted from Thompson (Part 3), *op. cit.*

A Reality Check—Estimating a Customer Waiting Cost from an Existing Labor or Service Standard

A simple reality test can be applied to the productivity or service standards in use at a hospitality business. The test operates by seeing what economic standard is implied by the productivity or service standards being used. For example, say that the service standard in use at a particular hospitality business forecast that three employees would be desirable in one planning period and four and five employees in later, separate planning periods. The manager could then analyze each of those periods to see what economic standards were consistent with those staffing levels. Taking the simplest economic standard, which uses a cost of customer waiting time, for example, the manager would find what range of customer waiting costs would result in three employees' being ideal in the first period, four employees' being ideal in the second period, and five employees' being ideal in the third period. A manager might find, for example, that the three employees needed in the first period would be consistent with customer waiting costs of $2.57 to $5.28 per hour. These cost ranges can then give the manager a sense of whether the original service standard was reasonable. With the costs of $2.57 to $5.28 per customer-hour, the manager might judge the service standard as reasonable if the service had a low cost of having customers wait (such as one with a low average transaction value, or one with few alternatives available to customers who thus pretty much would have to accept the wait). However, the manager would judge the original service standard as inappropriate if the service had a high cost of having customers wait (such as one with a high average transaction value, or one with many alternatives available to customers who thus would be reluctant to tolerate a wait). In this case, the original service standard should be replaced with one delivering a higher level of customer service.

An even better sense of the appropriateness of the original standard would be obtained by using the implied costs for all periods. So, for example, a manager could calculate the implied cost of customer waiting time in each of periods 1, 2, and 3 to judge the appropriateness of his productivity or service standard.

Employee requirements, or ideal staffing levels, are a key input to the task of developing a workforce schedule. One can translate demand forecasts into employee requirements in three ways, as described in the previous section of this report: using productivity standards, service standards, or economic standards. Productivity standards assume stable productivity factors, service standards apply consistent service levels, and economic standards are based on analysis of economic tradeoffs to determine employee requirements by job and by planning period.

Overview of Workforce Scheduling

This section identifies scheduling criteria, presents classic workforce-scheduling frameworks, identifies limitations of the classic frameworks, and presents contemporary workforce-scheduling frameworks.

Workforce-scheduling criteria. Workforce scheduling is usually performed in the context of either of two primary objectives. The first is to minimize the cost of the schedule, subject to constraints that ensure some level of staffing (e.g., a service standard). The second is to maximize the benefit to the organization of providing the correct schedule.

A common secondary objective of workforce scheduling is to maximize employees' satisfaction with the schedule. This necessitates the measurement of employee satisfaction, which often has many components. The goals of delivering a quality schedule (i.e., satisfying the primary objective) can conflict with the secondary objective of satisfying employee preferences. This issue is discussed further later in the report.

Classic workforce-scheduling frameworks. There are two classic conceptual frameworks for developing a workforce schedule. George Dantzig presented the first of these in 1954.[20] Dantzig's framework for workforce scheduling, or "D-Framework," is:

Objective: Minimize Schedule Cost

Subject to: For each period, the number of employees scheduled must equal or exceed the number of employees required

Simply put, D-Framework attempts to minimize the cost of providing the service while meeting or exceeding the employee requirements in every planning period. The number of employees needed in each period would be identified above in Translate. Dantzig's framework prohibits understaffing but does allow overstaffing. However, the objective of minimizing the schedule cost will eliminate overstaffing, if possible.

There are two key shortcomings of D-Framework, both related to the framework's prohibition of understaffing. First, a limited availability of employees may make it impossible to meet or exceed staffing requirements in all periods, in which case there would be no feasible schedule. In other words, D-Framework can break down when one attempts to implement it with real employees, that is, employees who commonly are available for work only at individually specified times.

A second shortcoming of D-Framework is my observation that managers often circumvent the understaffing prohibition when attempting to schedule employees using D-Framework. For example, if the managers develop a schedule that contains substantial overstaffing, they will reduce the employee requirements in some or all of the periods with no

[20] Dantzig, *op. cit.*

overstaffing and redevelop the schedule. Their hope is that the resultant schedule will have noticeably less overstaffing (which it often does). Though they have introduced understaffing when they reduced the employee requirements, the revised schedule frequently offers a better balance between the competing goals of eliminating both under- and overstaffing.

E.G. Keith presented the next new workforce-scheduling framework, K-Framework, in 1976.[21] In contrast to D-Framework, K-Framework allows understaffing. K-Framework is:

Objective: Minimize Schedule Cost, including the pseudo costs of under- and overstaffing

Subject to: For each period, the number of employees scheduled plus the employee shortage minus the employee surplus, equals the number of employees required

K-Framework develops a workforce schedule by attempting to minimize the labor-related costs of delivering the service, as well as pseudo costs (i.e., pseudo penalties) associated with under- and overstaffing. K-Framework uses tiered costs for under- and overstaffing, which tends to avoid large staff shortages or surpluses. A shortcoming of K-Framework is that extreme understaffing—where no employees are scheduled in a period—can occur. Needless to say, were it allowed to occur, such extreme understaffing would mean poor levels of customer service.

The academic literature on workforce scheduling has used D-Framework more than twice as often as K-Framework. This usage ratio is surprising, however, since most commercial workforce-scheduling packages implement some form of K-Framework and since earlier research has found that K-Framework, suitably modified to prohibit extreme understaffing, generally better satisfies the primary scheduling objectives identified above.[22]

Classic frameworks' limitations. Despite the general superiority of K-Framework over D-Framework, both classic frameworks have limitations that curtail their effectiveness. Both D-Framework and K-Framework suffer from what might be called a "single-period paradigm." A scheduling framework based on a single-period paradigm uses employee requirements that are set independently for each planning period. A single-period paradigm is problematic because the shifts worked by employees cover (i.e., affect) many planning periods.

In a service-standard environment, a single-period paradigm prohibits one from delivering a lower-than-ideal level of service in one period even though doing so might better control labor costs. The single-period paradigm makes it exceedingly difficult to determine employee-staffing levels that exactly provide the level of service for which one is striving. Another limitation of the classic frameworks in service-standard environments is their assumption that a surplus employee provides equal incremental, customer-service value, regardless of the period in which the surplus occurs. Such an assumption makes it impossible for either of the classic frameworks to maximize the level of service provided by a fixed labor cost or fixed labor hours.[23]

Similar limitations exist for D-Framework and K-Framework in an economic-standard environment. These

[21] Keith, *op. cit.*

[22] G.M. Thompson, "Representing Employee Requirements in Labor Tour Scheduling," *Omega*, Vol. 21, No. 6 (1993), pp. 657–671; Thompson (1995), *op. cit.*; and Thompson (1997), *op. cit.*

[23] For a detailed discussion of the shortcomings of the classic frameworks, see: Thompson (1997), *op. cit.*

frameworks' use of a single-period paradigm and their assumption that surplus employees are of equal cost regardless of the period in which the surplus occurs, means that the classic frameworks cannot provide the best schedule from an economic perspective.[24]

Contemporary scheduling frameworks. In the 1990s I presented two new conceptual frameworks for workforce scheduling. These two frameworks remedy the shortcomings found in the classical frameworks. One of these frameworks, the Contemporary Service Framework or CS-Framework, overcomes the limitations of the classic frameworks in service-standard environments,[25] and the other framework, the Contemporary Economic Framework or CE-Framework, overcomes the limitations of the classic frameworks in economic-standard environments.[26]

In simple terms, the CE-Framework for workforce scheduling is:

Objective: Maximize Total Schedule-Related Profit

Subject to: For each period, **(1)** ensure that a minimally acceptable number of employees are scheduled and **(2)** measure the number of additional staff providing an improvement in economic performance

The minimally acceptable number of employees in a period is typically less than the ideal number of employees for that period. CE-Framework measures the benefit of increasing the number of employees scheduled over the minimally acceptable level and finds the solution that best balances the monetary benefit of good service, the monetary cost of poor service, and the cost of delivering the service.

In contrast to the classic frameworks, which must include the ideal employee requirements as pre-specified inputs, CE-Framework and CS-Framework more closely link scheduling tasks two and three. CS-Framework and CE-Framework do this by recognizing that one cannot *a priori* (i.e., in TRANSLATE) determine the ideal employee requirements, because of the interdependence of staffing decisions across periods that exists in SCHEDULE. CS-Framework and CE-Framework thus exhibit what might be called a cross-period paradigm. Moreover, experiments have shown that CS-Framework and CE-Framework use the information provided by a cross-period paradigm to yield better schedules than the classic frameworks.[27] These experiments also showed that CS-Framework and CE-Framework perform better because of the way they handle information and not because they *require* better information. Unfortunately for hospitality firms, commercial workforce-scheduling systems currently use single-period paradigms or rudimentary forms of a cross-period paradigm.

Approaches and Techniques for Developing Workforce Schedules

This subsection first describes two approaches for generating schedules. It then describes two categories of techniques used in developing schedules. These two standard approaches for generating actual workforce schedules are a one-phase procedure and a two-phase procedure. In the following sections I explain why the one-phase approach is the better of the two and compare the two approaches on a simple problem.

Two-phase approach. The first phase of a two-phase approach develops a workforce schedule without regard to

[24] For a detailed discussion of these limitations, see: Thompson (1995), *op. cit.*

[25] Thompson (1997), *op. cit.*

[26] Thompson (1995), *op. cit.*

[27] *Ibid.*; and Thompson (1997), *op. cit.*

Exhibit 12

A comparison of one- and two-phase solutions for a simple scheduling problem

Hour	1	2	3	4	5	6	7	8
Employees needed	1	1	2	2	2	2	1	1
Employee A (V=Available)	V	V	V	V	V	V	V	
Employee B (V=Available)			V	V	V	V	V	V

Shift-based schedule (two phase)

Shift 1 (W=Work)	W	W	W	W	W	W	W	W
Shift 2 (W=Work)			W	W	W	W		
Total scheduled	1	1	2	2	2	2	1	1
Net staffing level	0	0	0	0	0	0	0	0

Holistic schedule (one phase)

Shift 1 (W=Work)	W	W	W	W	W	W		
Shift 2 (W=Work)			W	W	W	W	W	W
Total scheduled	1	1	2	2	2	2	1	1
Net staffing level	0	0	0	0	0	0	0	0

As explained in more detail in the text on the next page, forecasts of customer activity show that during the above eight-hour period the establishment will need one employee on duty during the first two and last two hours. Assume that shifts must be at least four hours and at most eight hours long and that only the two employees shown are available for work. To handle a mid-period rush, two employees should be on duty, as shown in the "employees needed" row. As shown in the top, shift-based schedule, scheduling without regard to employee availability can put one employee on duty for the entire eight hours and put a second employee on during the four-hour rush. The schedule meets customer-service standards by balancing the number of employees needed with those scheduled, as indicated by the zeros in the "net staffing level" line. Clearly this schedule is unacceptable, however, since neither employee could cover the eight-hour shift. In contrast, the bottom, holistic schedule, accounts for the employees' availability and balances the workload (and paychecks) for both, while still meeting the customer-service standards.

employees' availability or preferences, while the second phase assigns shifts to specific employees. Although two-phase approaches generate schedules relatively quickly, they generally yield inferior schedules than do one-phase approaches, for two reasons. First, two-phase approaches often yield poorer matches between the number of employees scheduled and the number of employees needed. Second, shift-based approaches may not always satisfy certain restrictions, such as on employee minimum hours.

Holistic. A one-phase, or holistic, approach to developing workforce schedules considers employee information in the process of developing the shift schedule. It thus overcomes the limitations of a two-phase approach. However, holistic approaches can be slow unless implemented well. Existing commercial workforce-scheduling systems use both one- and two-phase approaches, though the one-phase approach is becoming more common due to the problems associated with two-phase approaches.

Here is a sample scheduling problem that illustrates the difficulty that can arise with a two-phase approach. Exhibit 12 shows the employee requirements in each of eight hour-long planning periods. Assume that shifts must be at least four hours long and no more than eight hours long. Finally, assume that the following two employees are available for work: Employee A is available from hour 1 to hour 7, while Employee B is available from hour 3 to hour 8. For simplicity of presentation, the shift-based and holistic approaches will be compared using D-Framework.

As Exhibit 12 shows, the shift-based schedule perfectly matches employee requirements in all periods—that is, it has no under- or overstaffing. The shift-based schedule has two shifts: the first covering hours 1 through 8, and the second covering hours 3 through 6. The problem with this schedule is obvious when one attempts to perform the second phase—assigning the shifts to the employees. Either employee could work Shift 2, but neither employee could work Shift 1. A shift-based approach would then have to **(1)** leave Shift 1 unassigned, which would result in substantial understaffing, or **(2)** modify Shift 1 so that an employee could cover at least a portion of the shift. However, any such modification would be problematic in any situation but a simple one like this. Instead, it makes more sense to factor employee availability in from the start, as is done by a holistic approach. A holistic approach to this problem develops a schedule that has no under- or overstaffing and which has two shifts. The first shift, covering hours 1 through 6, would be assigned to Employee A, and the second shift, covering hours 3 through 8, would be assigned to Employee B. By considering employee availability while developing the schedule, a holistic approach will not schedule shifts that cannot be staffed.

Techniques for solving workforce-scheduling problems. The many methods, or algorithms, for actually developing workforce schedules can be categorized as either optimal or heuristic. An optimal procedure is one that is guaranteed to find the best possible schedule. Unfortunately, the types of problems where one can find an optimal schedule are usually much less complex than those occurring in real hospitality firms. Even if optimal procedures are applied at both phases of a two-phase approach, the overall solution will usually not be optimal because a two-phase approach does not consider the entire scheduling problem at one time.

As explained in the box on the next page, a heuristic procedure is meant to obtain a good schedule quickly. Heuristic procedures may find the optimal schedule (quite by accident), but have no means of verifying that the schedule is, in fact, optimal. Developing an effective heuristic is an art. One strives to develop a solution procedure that incorporates relevant information and that is effective across a broad range of scheduling scenarios. Optimally developing one-phase schedules is impractical at this point for most real scheduling situations due to the extreme time required. Consequently, most commercial workforce-scheduling systems are heuristic based.

Scheduling Considerations

This section discusses the considerations in schedule development: namely, controllable and uncontrollable work, scheduling flexibility, employee issues, hard and soft constraints, forced and voluntary overtime, and scheduling horizon duration.

Controllable and uncontrollable work. As discussed earlier in this report, uncontrollable work is work over which there is little temporal control—specifically, the customer service that must be performed when the customers are in the service system. Controllable work, again, is work over which there is

Exhibit 13

Comparison of methods for scheduling controllable work

		Hour	1	2	3	4	5
		Ideal number of staff to perform the uncontrollable work	4	3	5	3	3
Pre-Assigned Controllable Work		**Pre-assigned workload of controllable work**		1		1	1
		Total employee requirements	4	4	5	4	4
	Optimal D-Framework Solution	Shift 1	w	w	w		
		Shift 2	w	w	w		
		Shift 3	w	w	w		
		Shift 4	w	w	w		
		Shift 5			w	w	w
		Shift 6			w	w	w
		Shift 7			w	w	w
		Shift 8			w	w	w
		Total scheduled employees	4	4	8	4	4
		Overstaffing	0	0	3	0	0
Scheduled Controllable Work	*Optimal D-Framework Solution*	Shift 1	w	w	w		
		Shift 2	w	w	w		
		Shift 3	w	w	w		
		Shift 4	w	w	w		
		Shift 5			w	w	w
		Shift 6			w	w	w
		Shift 7			w	w	w
		Total scheduled employees	4	4	7	3	3
		Scheduled controllable work		1	2		
		Total workload	4	4	7	3	3
		Overstaffing	0	0	0	0	0

Note: The above example assumes that the employees are properly cross-trained to handle all the tasks at hand. The scheduling situation also assumes that the employee shifts are only three hours long and, thus, can only start in hours one, two, or three.

some degree of temporal control. Controllable work is useful from a scheduling perspective, simply because it is controllable. The ability to schedule the work, within some limits, offers a modicum of flexibility.

Controllable work should be scheduled in conjunction with the schedule for uncontrollable work.[28] Doing so enables the workforce to operate at a higher utilization than can be achieved when one schedules the controllable work only to fill periods of idle time. Here is an example. Consider scheduling cross-trained employees to cover controllable and uncontrollable work in a five-hour scheduling horizon. Exhibit 13 presents the ideal number of employees to perform the uncontrollable work. Three hours of controllable work must be performed. This work can all be done in any single hour, or it can be spread across two or more of the five hours. Shifts are three hours long and can start in hours 1 through 3.

[28] Thompson (1992), *op. cit.*

Heuristics

A heuristic is a logic-based procedure designed to yield good schedules quickly. Consider the task of trying to construct a workforce schedule. One must start with a blank slate—no shifts—and add shifts until there are enough. How should one select shifts to add to the schedule? In an earlier paper I evaluated a set of 20 rules for constructing schedules.* These 20 rules were based on five criteria, each used as a primary criterion and as a tie-breaker for the others. These five criteria were to add the shift that: **(1)** covers the period having the greatest single-period staff shortage; **(2)** covers the highest average short staffing; **(3)** offers the greatest reduction in schedule cost per working period; **(4)** offers the greatest improvement in schedule smoothness (where smoothness is measured as the absolute difference in net staffing levels from period to period); and **(5)** covers the periods having the highest average ratio of the number of employees still needed in a period divided by the number of still-unassigned employees who could work the period. The following example shows how two of these rules would work. Two shifts, each one four hours long, are being considered for addition to the schedule. The first shift covers periods where the net staffing levels are 0, -1, -2, and 0 (i.e., workers are still needed only in the second and third hours of the shift). The second shift covers periods where the net staffing levels are -1, -1, -1, and -1. The criterion that adds the shift covering the maximum staff shortage would select the first of these shifts, while the criterion that adds the shift covering the greatest average short staffing would select the second of these shifts.

*G.M. Thompson, "A Simulated-annealing Heuristic for Shift Scheduling Using Non-continuously Available Employees," *Computers and Operations Research*, Vol. 23, No. 3 (1996), pp. 275–288.

If one were to assign the controllable work prior to developing the schedule of uncontrollable work, one would be tempted to assign the controllable work to the periods where uncontrollable work was the lowest, yielding the total workload requirements shown in Exhibit 13. Using D-Framework for simplicity of presentation, eight shifts make up the best schedule, four commencing in hour 1 and four commencing in hour 3. This schedule has three hours of idle time, all occurring in hour 3.

When one schedules the controllable work simultaneously with shifts, on the other hand, the best D-Framework schedule comprises only seven shifts, four commencing in hour 1 and three commencing in hour 3. One hour of controllable work is assigned to hour 2 and two hours are assigned to hour 3. This schedule has no scheduled idle time; the total of 21 hours of scheduled time equals the total required workload.

Comparing the two schedules, one sees the importance of taking advantage of the flexibility that controllable work offers. This flexibility yields lower staff shortages and surpluses; in other words, flexibility can allow one to achieve the primary scheduling objectives more effectively. The next section introduces other forms of scheduling flexibility.

Scheduling flexibility. Flexibility can be a boon that enables the scheduler more closely to match the actual number of employees scheduled with the ideal number of employees. Such flexibility can also be a bane, however, in that it can greatly increase the complexity of the scheduling situation and often will cause solution procedures to work more slowly.

Flexibility options include controllable work, as noted above, variable shift lengths (including overtime), break timing, alternate start times, and cross-utilization of employees in different jobs within and across shifts. Many academic studies have been conducted on the value of flexibility. As one might expect, the benefits of increasing flexibility depend on the degree of flexibility already present. Low flexibility environments will benefit most from increasing flexibility; environments with extensive flexibility, on the other hand, will benefit much less from increasing flexibility.

Employee considerations. The many employee characteristics that can be considered during schedule development are grouped into environmental and preferential characteristics. Environmental characteristics include seniority,[29] the skills the employees have to perform the different positions being scheduled, days-off considerations (that is, whether consecutive days off must be assigned), and restrictions on the minimum and maximum daily hours and weekly hours and the times at which employees are not available for work. The environmental considerations, though employee-related, are generally outside the employees' control, at least over the duration of the scheduling horizon.

Employees' preferential characteristics should include preferences for the total daily and weekly work hours, for days off, task assignments, and work positions (e.g., a particular station in a full-service restaurant; the fryer or the counter in a QSR). Employees may also express preferences for the times at which they could work but would rather not do so, when they ideally would like to start and finish work, and the length of meal break they would like.

[29] G.M. Thompson, "Assigning Telephone Operators to Shifts at New Brunswick Telephone Company," *Interfaces*, Vol. 13, No. 4 (July–August 1997), pp. 1–11.

Needless to say, both the environmental and preferential characteristics can vary from employee to employee. For example, a particular employee's highest preference might be getting her desired total work hours, another's highest preference might be to get the work station he most prefers, and a third employee just wants Thursday night free. One should not view employee preferences as reducing management's ability to deliver the service. This is because employees rarely have the same preferences for schedule characteristics and any complementary differences in preferences can be exploited when developing a schedule.

Hard and soft constraints. Constraints are the factors that limit one's ability to develop a workforce schedule that best satisfies the primary objective (i.e., a good schedule, from the organization's perspective) and the secondary objective (which is a schedule where employees' preferences are all perfectly satisfied). The range of restrictions in real-world workforce-scheduling problems often results in conflicting constraints. Hard constraints are those that must be satisfied in a schedule, if at all possible. Soft constraints, in contrast, are those that should be satisfied in a schedule when other factors permit. In other words, hard constraints should not be violated (except perhaps by another hard constraint), while soft constraints can be violated. In general, hard constraints will relate to the environmental characteristics identified earlier; soft constraints are often associated with preferential characteristics. Thus, hard constraints will often include the restrictions that define the minimum and maximum acceptable daily and weekly hours for employees, that ensure employees are scheduled only in jobs for which they are skilled, and ensure that employees are never scheduled at times they are unavailable. Soft constraints would include attempting to schedule employees to their ideal daily and weekly hours, for example.

In this regard, it may be helpful to think about a hierarchy of constraints. Constraints higher in the hierarchy would take precedence over constraints lower in the hierarchy. It is also often useful to place small-scope constraints higher in the hierarchy than the large-scope constraints. For example, the minimum and maximum daily and weekly work-hour limits would all be hard constraints. However, the restrictions on daily work hours would have a higher priority than those on weekly work hours. This would allow one to resolve the situation of an employee who is available five days in the week and who could work a maximum of seven hours per day being required to work at least 38 hours for the week. The hierarchy would satisfy the daily work hours at the expense of violating the weekly work hours.

Once demand is forecasted and the schedule is planned, managers must take into account the actual situation and make necessary adjustments.

The reason that the limits on daily and weekly work hours constitute hard constraints is that firms often have contractual or obligatory relationships with employees that define the amount of work hours the employees should receive. I once viewed a scheduling system in use by a hospitality company that fails to treat employees' minimum weekly hours as a hard constraint. Managers using this system spent a considerable amount of "tinker time" adjusting the schedule manually to ensure employees were getting their specified weekly minimum hours.

Forced and voluntary overtime. Hospitality businesses often operate in high-demand situations where the employees cannot cover all the required workload when working regular hours. It is useful to consider how to handle such situations from a scheduling perspective. One option is to schedule forced overtime. Employees can be assigned longer daily shifts or an additional work day to yield higher total weekly hours. A second option is to schedule voluntary overtime hours or shifts. Forcing overtime is straightforward, providing that one's scheduling system allows it. Scheduling voluntary overtime, though, is less straightforward and merits further consideration.

One approach to handling voluntary overtime is to develop unassigned shifts, that is, shifts that are needed but are not assigned to any particular employee. These shifts can then be posted and employees can sign up for shifts that they are willing to work. A second approach is to develop the schedule by assigning the overtime to specific employees using information as to which employees would be willing to work overtime and whether they would be willing to pick up an extra shift, work longer shifts, or both. Of these two alternatives for handling voluntary overtime, the second is preferable. The reason relates to the difference between the schedule as originally developed, including unassigned shifts, and the actual schedule—the schedule that actually is worked (where only a subset of the unassigned shifts may actually be picked up by employees). An original schedule might be good at delivering employees when needed. However, there is no guarantee that the unassigned shifts will be scheduled at times that allow employees to work them. If the whole schedule is optimized, including the unassigned shifts, then if any unassigned shift cannot be assigned, the resultant schedule will not be as good as it could be. Scheduling voluntary overtime for specific employees overcomes the problems associated with scheduling unassigned shifts.

Scheduling-horizon duration. Workforce schedules in the hospitality industry are commonly developed for one- to two-week periods, though sometimes the schedules are developed only quarterly. Firms that use long scheduling horizons unfortunately lose the ability to respond well to changes in demand. Accurately forecasting a week to ten days ahead is simple compared to accurately forecasting three months ahead. I have seen a hospitality firm get trapped with a long scheduling horizon because it used a two-phase approach to develop its workforce schedule. Employees of the firm selected their schedules according to their seniority and productivity, a process so cumbersome that the firm could justify performing it only once per quarter. A better solution would have been for the firm to use a one-phase solution, which incorporates employee preferences and uses these preferences when developing a schedule. It could then reduce its scheduling horizon to a much more effective one- to two-week period.

Summary

The tasks described in this and the two preceding sections have focused on planning, which is the process by which a schedule is developed for the demand that is anticipated. In the next section of the report, I focus on the final task in workforce scheduling—controlling the schedule in real time. Real-time control ensures that the actual schedule worked by employees meets the expectations of the planned schedule.

Section 5

Controlling the Schedule in Real Time (Control) [30]

After completing the three steps covered so far in this report—Forecast, Translate, and Schedule—a manager would have a forecast of the elements of the service transaction (particularly, customer arrival rates), a list of the number and skills of employees needed, and a specification of who is working where and when. The first three tasks are all planning activities that are conducted in advance of the service transactions. In contrast to those three tasks, the final task, Control, involves the real-time control of the schedule, in which the manager assesses whether the schedule is ensuring that customers are actually being served as planned.

Control, which involves comparing operating reality to the planned schedule, is the essential final piece to ensuring that your customers will be served appropriately. The difficulty in making sure service is as it should be occurs when real-time imbalances between labor capacity and customer demand arise. Such imbalances occur because demand rarely materializes exactly the way one has forecast it and because employees do not always perform the way one anticipates (e.g., they may be sick or arrive late). The uncertainty in demand forecasts and employee performance highlights the need for effective real-time control that

[30] Adapted from Thompson (Part 4), *op. cit*.

ensures that the *actual* schedule is effective.

In this section of the report I explain how a manager can assess with reasonable certainty whether the forecasted schedule is, in fact, matching the day's customer demand. I'll also touch on some of the actions a manager can take when demand does not match the forecasted schedule. In particular, I'll explain the value of my earlier recommendation of having available cross-trained employees, particularly in situations when demand is uncertain.

Because the approach I outline works best when customer counts are relatively high, the material in this section will have the greatest applicability in high-volume restaurants and large hotels with substantial walk-in demand, as well as such other high-volume operations as reservation centers. The relatively high variability that occurs with low customer counts reduces one's ability to make an early prediction of the day's likely business volume. Even though most of my analysis is aimed at high-volume operations, hospitality services with low customer counts can use some of the techniques I mention, in particular those that are short lived, as I explain next.

Real-time-control Actions

Real-time-control actions can be categorized either as short lived or long lived. Short-lived actions are those that affect only a small period of the operating day, typically a few minutes to an hour. Such actions are easily revocable. Short-lived actions include sending employees to or recalling them from break, extending the length of an employee's shift (including overtime), and reassigning employees to different jobs. Long-lived actions are those that will affect a period longer than an hour and entail a greater commitment of resources. They include sending employees home early, calling additional employees in to work, and reassigning employees to different jobs.

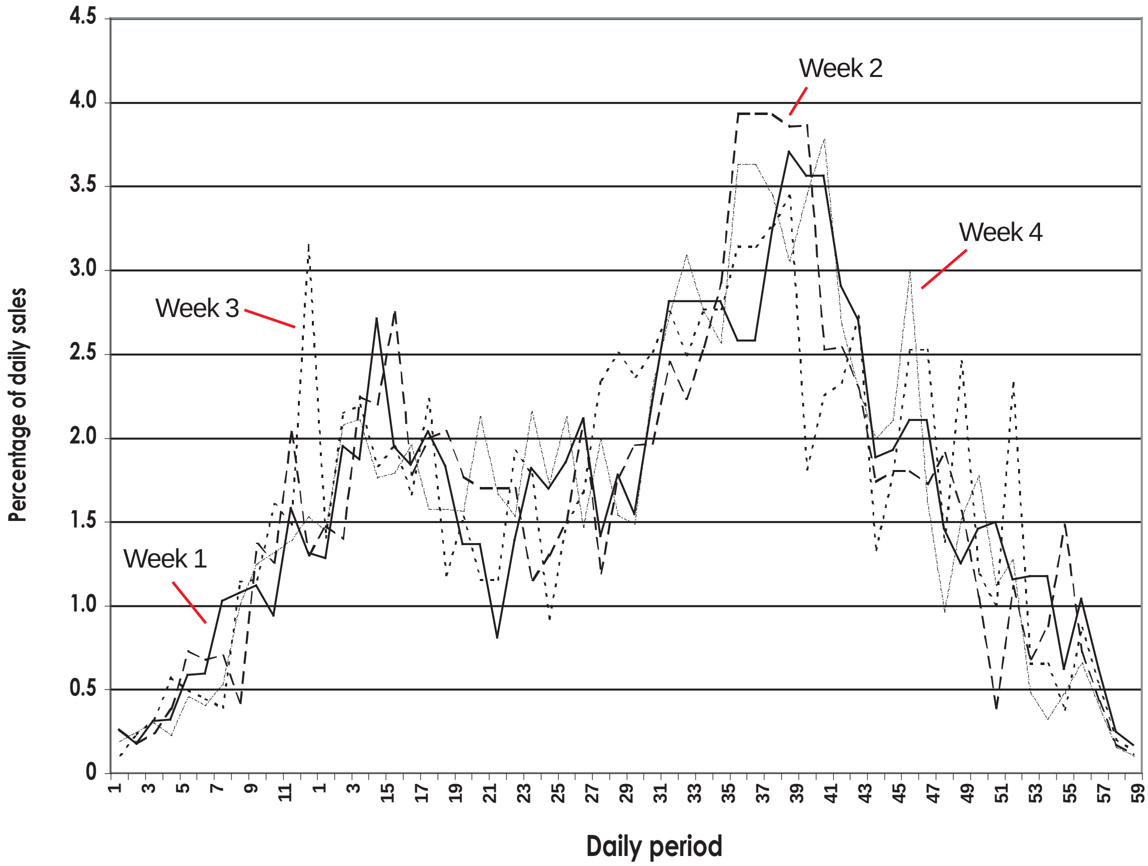
The key issues of real-time control are determining when to take an action that modifies the original schedule and (if that determination is positive) whether to take a short-lived or long-lived action. Short-lived actions have a relatively small effect on costs and on customer service, while long-lived actions not only affect operations, but they can be difficult to reverse. Thus, for a manager to confidently take a long-lived action, she must be able to predict the hospitality operation's demand for that day. Say that demand on Mondays is fairly consistent, for example. With that consistency, a manager can make a statement like: "If we're slower than we anticipated by 11 o'clock, then it's likely we'll be slow for the whole day." The final step in scheduling is to be able to quantify that statement and to identify as early as possible whether a given day *as a whole* will be slower or busier than was forecast. A manager who can make that judgment can confidently take long-lived actions, such as sending employees home. Without that predictive confidence, however, the manager will have at her discretion only short-lived actions (e.g., assigning more employees to side work).

A Step-wise Approach to Tracking Demand

This section of the report primarily examines a five-step process for predicting a day's customer counts. The steps are as follows: determine whether the operation enjoys consistent demand; identify the proportion of sales accruing to each planning period; categorize each day by its business volume; run a simulation of each day's business pattern to develop business-volume-consistency charts; and track customer counts

Exhibit 14

Same-day sales comparisons, week to week



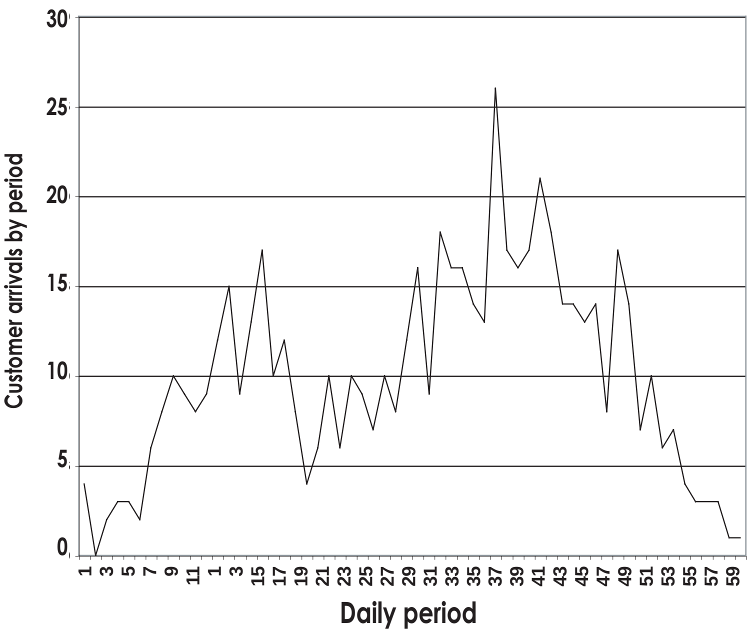
Sales are shown for every 15-minute period as a proportion of total daily sales for a particular day of the week (e.g., Mondays) for four consecutive weeks.

against the simulation to predict day-end business volume. I'll explain each of these steps, although some steps will be familiar from the initial process of developing the demand forecast.

Step 1—**Determine the extent to which each day has a consistent demand pattern.** I addressed the issue of consistency in within-day demand earlier in this report. Consistent within-day demand means that each period within the day has a consistent portion of the daily demand, regardless of the total volume of business on that day. Exhibit 14 provides an example of this consistency. In this step, the manager plots the sales in each planning period as a percentage of sales for the day. The graph shows that demand is similar on the four consecutive Mondays, building to a secondary peak around period 15, experiencing a lull through period 25, and then building to the primary peak in period 40, followed by a tapering off throughout the remainder of the day. Running a correlation of the daily data

Exhibit 15

Period-by-period realization of a day's customer demand (hypothetical)



This graph represents a hypothetical period-by-period realization of a level-1 (light volume) day. A key feature of the measurement is the variability, as represented by the sharp peaks and valleys from period to period, which represent the randomness of customer demand.

yields strong correlations of 0.79 and higher, which indicates consistent within-day demand.

Step 2—**Identify the proportion of daily sales occurring in each planning period.** Once a manager has established whether within-day demand is consistent, she needs to identify the proportion of daily sales that occur in each planning period (any given division of the day, but often a 15-minute period). One begins by calculating an average proportion of sales in each period and applying the smoothing technique I described earlier to obtain the demand curve illustrated in Exhibit 5 (page 22). As discussed earlier, managers should be able to articulate the reasons customer demand materializes at the times it does. For example, given the nature of the service, there are reasons why the peaks and valleys in demand fall at the times they do.

Step 3—**Categorize each day according to its business volume.** The next step is to label a given day according to a customer-volume category. A reasonable way to do this is to establish, say, five categories of business volume that cover the range of the operation's total daily customer counts. Level 1 would be the lowest volume, while level 5 would be the highest.

For example, consider a hospitality service that typically serves 1,000 patrons on an average day, but where customer counts can range from 500 to 1,500. One could set up demand categories by dividing the thousand-customer spread into five even levels. Under that scheme, the five customer-volume levels would be 500–700 (level 1), 701–900, 901–1,100, 1,101–1,300 and 1,301–1,500 (level 5).

With those categories in mind, a manager would tally the actual customer demand by planning period over the course of the day. Such a graph is termed a realization. Exhibit 15 shows a hypothetical period-by-period realization of a level-1 day, during which a total of 598 customers are served. Note that the demand realization is generally consistent with the average business by period shown in Exhibit 5. There's a peak around period 15, a lull until approximately period 25, and then the facility hits its greatest demand around period 40. However, a key feature of Exhibit 15 is its variability, as represented by the sharp peaks and valleys from period to period. This variability—the randomness of customer demand—is the characteristic of service systems that is the prime driver of the need to make real-time capacity adjustments.

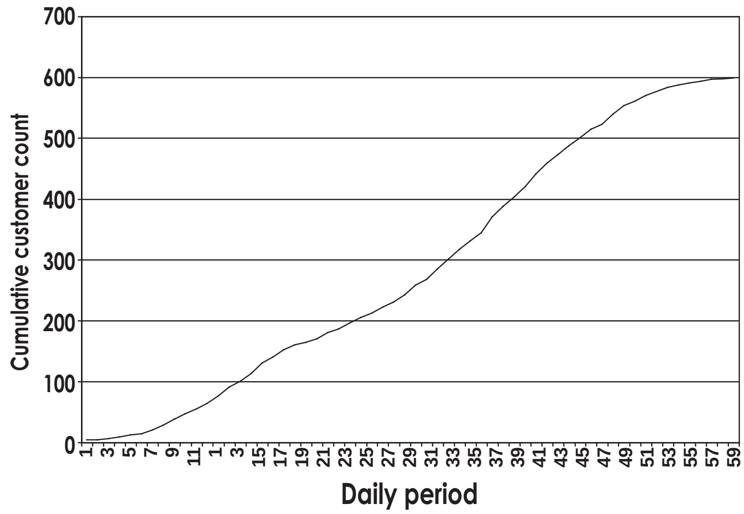
A more useful tool than the period-by-period realization shown in Exhibit 15, however, is a cumulative-business-volume realization for the same data, which is found in Exhibit 16. This graph shows that 100 customers were served by period 15; 200 customers were served by period 26; and 300 customers were served by period 34. A manager needs this cumulative realization for the next step of the process, which is to assess the extent to which the realized demand is consistent with the forecasted demand.

Step 4—**Simulate realizations of the business-volume categories and develop business-volume-consistency charts.** To assess the fit between actual demand and forecasted demand, one needs a set of cumulative realizations against which to compare a given day's demand pattern. To get that set of realizations, the manager must either collect real data or simulate a set of cumulative realizations of these business volumes. One should collect or simulate over 100 realizations of each business volume (that is, 100 days of demand figures). Once these realizations are collected or simulated, one can develop business-volume-consistency charts that show the range of cumulative customer counts by period within a day. Because of the difficulty of collecting enough real data to develop the business-volume-consistency charts, I recommend that you develop a simulation.

Simulation is a useful tool for generating more data about operations than are readily available. Customer arrivals in the hospitality business typically follow Poisson distributions (a calculation useful in situations like this where the only information is an average rate of a discrete event's occurrence).[31] With a Poisson distribution, one can develop an equation that specifies the typical arrival patterns. Then, using a series of random numbers, one can simulate customers' arrival times. Doing this one time is not especially useful, but running the calculation through many iterations gives one a reasonable simulation of arrivals for different business volumes.

Exhibit 16

Cumulative realization of a day's hypothetical customer demand

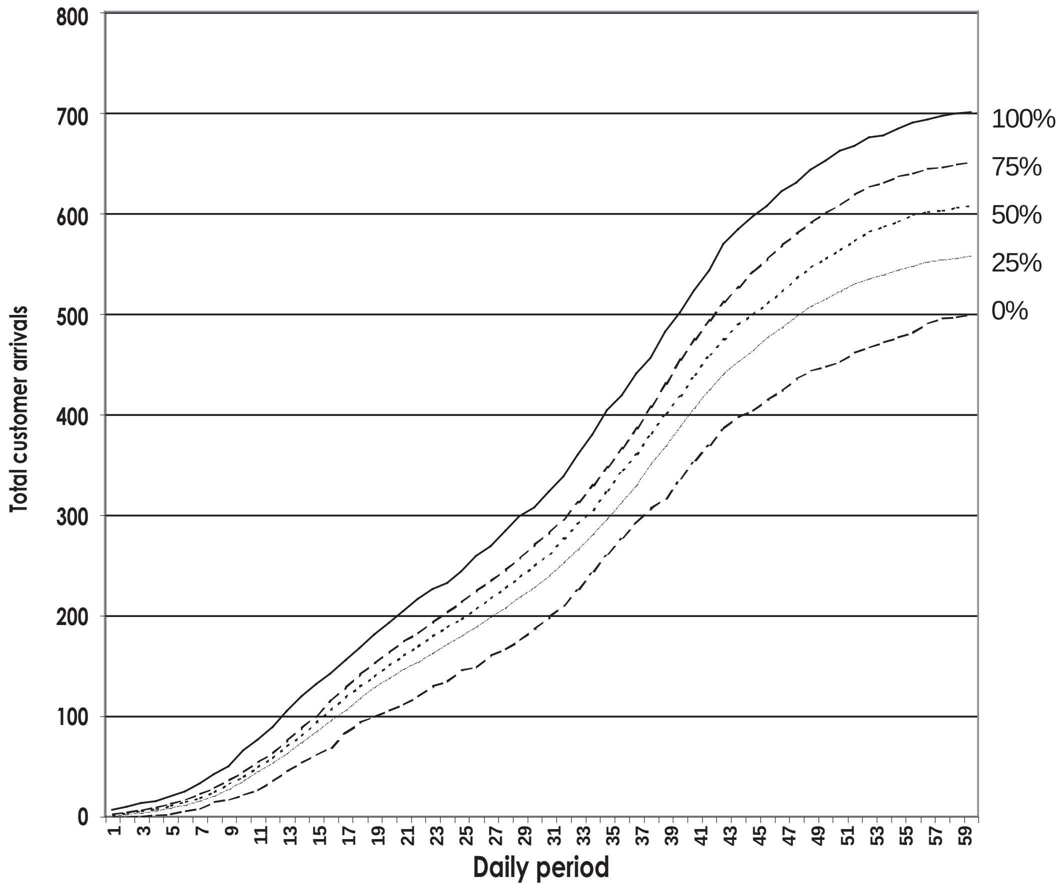


This graph shows the cumulative realization for the customer-arrival data in Exhibit 15. A manager needs the cumulative realization to assess the extent to which the realized demand is consistent with the forecasted demand.

[31] Poisson's equation provides a way to estimate the distribution of discrete, repetitive events. For a definition, see: en.wikipedia.org/wiki/Poisson_distribution.

Exhibit 17

Business-volume-consistency chart (low-demand day)

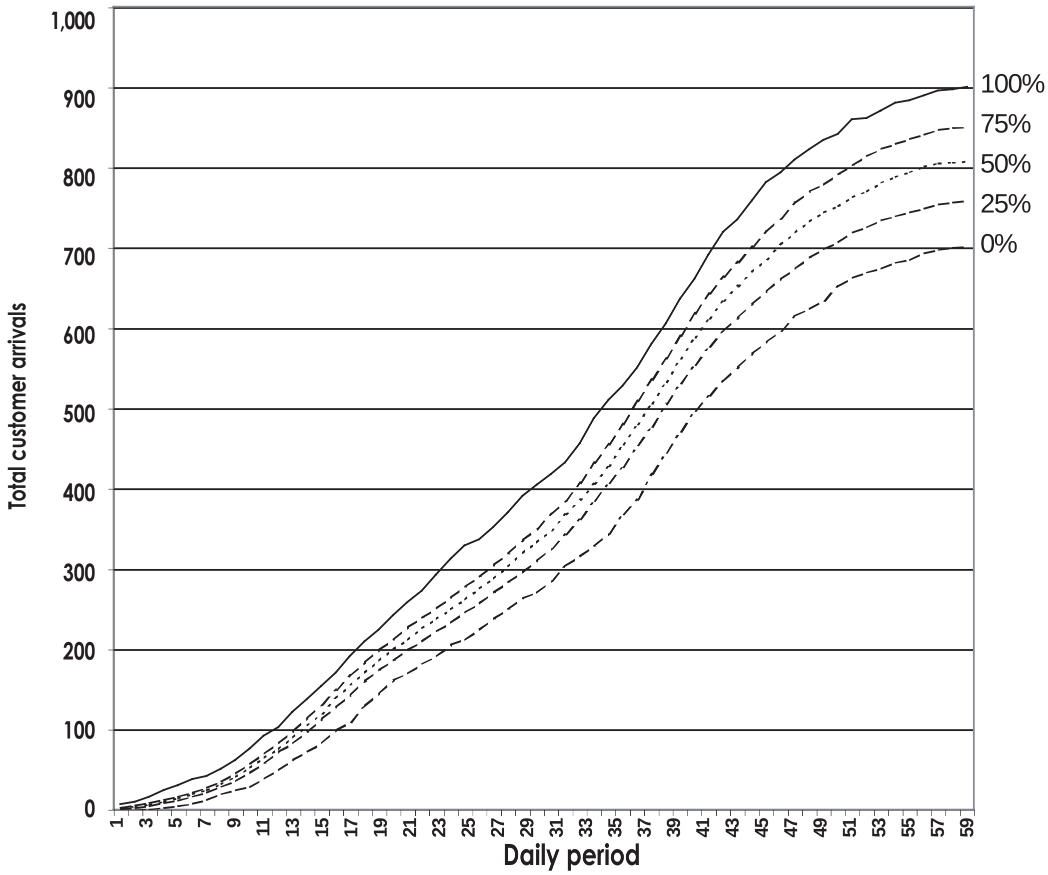


As explained in the text, Exhibit 17 shows level-1 volume (a total daily customer count in the 500-700 range) based on 200 simulated realizations. The 100-percent (topmost) line identifies the greatest number of customers served at any given point for days of a particular demand level while the zero-percent (bottommost) line indicates the fewest customers served at any point.

Examples of business-volume-consistency charts are shown in Exhibits 17 and 18. Here's how they help you determine whether your schedule forecast is holding. Exhibit 17 shows level-1 volume (500 to 700 customers served for the day) based on 200 simulated realizations. The 100-percent line identifies the greatest number of customers served at any given point for days of a particular demand level (while the zero line indicates the fewest customers served at any point). The 50-percent line is the median number of customers served by that point, for a given day's demand. The 25-percent line is the first quartile and the 75-percent line is the third quartile of the customer counts. Looking at period 30, and the 100-percent line, one will note that on any day that this operation served more than 307 customers by period 30, the operation *never* recorded a level-1 day. In other words, demand at that level by that time foretells more customers coming so that the day's demand will exceed 700. From the zero line we observe, by the same token, that on no day did we serve fewer than 188 customers by period 30 and still end the day in

Exhibit 18

Business-volume-consistency chart (medium-demand day)



This chart represents a level-2 day (a day with a total customer count in the 701-900 range).

the level-1 business volume. Finally, the 50-percent line shows that 50 percent of the time, the customer count was 250 or lower by period 30. (In other words, the 50-percent line is the median.)

The business-volume-consistency charts operate in similar fashion for other levels of business. Exhibit 18, for example, is a business-volume-consistency chart for a level-2 day (total demand of 701–900). Again keying on period 30, observe that the operation never had served fewer than 272 or more than 405 customers at that point on a level-2 day, while 50 percent of the time it served 334 or fewer customers by period 30.

Although it is not readily apparent from Exhibits 17 and 18, the business-volume-consistency charts are not always distinct for separate volume levels. For example, consider a situation in which the operation had served 290 customers by period 30. This customer volume is within the range the operation experienced for level-1 days *and* within the observed range for level-2 days. A manager must be able to discern those differences to complete the process by predicting the day-end customer count.

Exhibit 19

Sample use of customer counts to predict daily volume

Period	Cumulative actual customer count	Number of comparable cases (% of total comparable cases) Level 1	Level 2	Level 3	Level 4	Level 5	Total comparable realizations
1	2	68 (25.56)	69 (25.94)	51 (19.17)	43 (16.17)	35 (13.16)	266
2	7	185 (23.99)	179 (23.22)	150 (19.46)	145 (18.81)	112 (14.53)	771
3	10	71 (16.55)	100 (23.31)	94 (21.91)	94 (21.91)	70 (16.32)	429
4	13	63 (18.16)	96 (27.67)	78 (22.48)	74 (21.33)	36 (10.37)	347
5	14	39 (31.97)	46 (37.7)	21 (17.21)	12 (9.84)	4 (3.28)	122
6	21	117 (35.67)	119 (36.28)	62 (18.9)	26 (7.93)	4 (1.22)	328
7	25	63 (35.80)	74 (42.05)	33 (18.75)	6 (3.41)	0 (0.00)	176
8	28	52 (61.90)	28 (33.33)	4 (4.76)	0 (0.00)	0 (0.00)	84
9	37	115 (69.28)	48 (28.92)	3 (1.81)	0 (0.00)	0 (0.00)	166
10	45	99 (75.00)	30 (22.73)	3 (2.27)	0 (0.00)	0 (0.00)	132
11	53	93 (83.78)	18 (16.22)	0 (0.00)	0 (0.00)	0 (0.00)	111
12	65	119 (87.50)	17 (12.5)	0 (0.00)	0 (0.00)	0 (0.00)	136
13	73	74 (92.50)	6 (7.5)	0 (0.00)	0 (0.00)	0 (0.00)	80
14	87	102 (92.73)	8 (7.27)	0 (0.00)	0 (0.00)	0 (0.00)	110
15	103	101 (89.38)	12 (10.62)	0 (0.00)	0 (0.00)	0 (0.00)	113
16	116	73 (91.25)	7 (8.75)	0 (0.00)	0 (0.00)	0 (0.00)	80
17	128	62 (95.38)	3 (4.62)	0 (0.00)	0 (0.00)	0 (0.00)	65
18	147	82 (90.11)	9 (9.89)	0 (0.00)	0 (0.00)	0 (0.00)	91
19	155	33 (89.19)	4 (10.81)	0 (0.00)	0 (0.00)	0 (0.00)	37
20	171	54 (85.71)	9 (14.29)	0 (0.00)	0 (0.00)	0 (0.00)	63
21	186	38 (74.51)	13 (25.49)	0 (0.00)	0 (0.00)	0 (0.00)	51
22	198	32 (68.09)	15 (31.91)	0 (0.00)	0 (0.00)	0 (0.00)	47
23	204	23 (85.19)	4 (14.81)	0 (0.00)	0 (0.00)	0 (0.00)	27
24	215	33 (86.84)	5 (13.16)	0 (0.00)	0 (0.00)	0 (0.00)	38
25	221	14 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	14
26	229	22 (95.65)	1 (4.35)	0 (0.00)	0 (0.00)	0 (0.00)	23
27	239	33 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	33
28	248	27 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	27
29	259	30 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	30
30	270	27 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	27
31	288	42 (97.67)	1 (2.33)	0 (0.00)	0 (0.00)	0 (0.00)	43
32	307	41 (93.18)	3 (6.82)	0 (0.00)	0 (0.00)	0 (0.00)	44
33	332	48 (92.31)	4 (7.69)	0 (0.00)	0 (0.00)	0 (0.00)	52
34	341	21 (95.45)	1 (4.55)	0 (0.00)	0 (0.00)	0 (0.00)	22
35	352	23 (92)	2 (8)	0 (0.00)	0 (0.00)	0 (0.00)	25
36	368	35 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	35
37	387	40 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	40
38	412	50 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	50
39	443	53 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	53
40	464	38 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	38
41	485	31 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	31
42	498	21 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	21
43	513	24 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	24
44	531	27 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	27
45	543	15 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	15
46	550	8 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	8
47	557	10 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	10
48	571	20 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	20
49	576	7 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	7
50	582	6 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	6
51	594	12 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	12
52	604	11 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	11
53	607	3 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	3
54	615	7 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	7
55	620	5 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	5
56	624	3 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	3
57	628	3 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	3
58	633	3 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	3
59	636	5 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	5
60	636	3 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	3

Step 5—**Track customer counts and predict day-end business volume.** Given the overlap between the business-volume-realization curves, one might question how a manager could hope to distinguish, say, a level-1 day from a level-2 day. Here's one way to do that. The idea is to match a given day's customer counts with the appropriate consistency curve. One does this by recording a cumulative customer count early in the day and checking how the resulting graph predicts the day will end up based on that count. As one tracks the cumulative customer count for a given day, one can compare the actual cumulative demand to the simulated-realization curves. Make a count of the total number of comparable cases in the simulated realizations and record the frequency with which each business volume contributed to the total comparable realizations. A comparable case in this instance is one where the simulated cumulative customer count in the previous period was equal to or less than the current customer count while at the same time the simulated customer count in the current period equals or exceeds the current customer count. For example, let's say we had served 14 customers by the end of the fifth period of the day. Two comparable realizations would be one in which 14 or fewer customers had been served by end of period 4, and one where 14 or more customers had been served by the end of period 5.

As one moves through a day, one can monitor the level of business experienced to that point and compare it to the simulated realizations under different business volumes. Exhibit 19 shows an example of this approach. By the end of period 10, 45 customers had arrived on the particular day being tracked. Of the 200 realizations for level-1 business

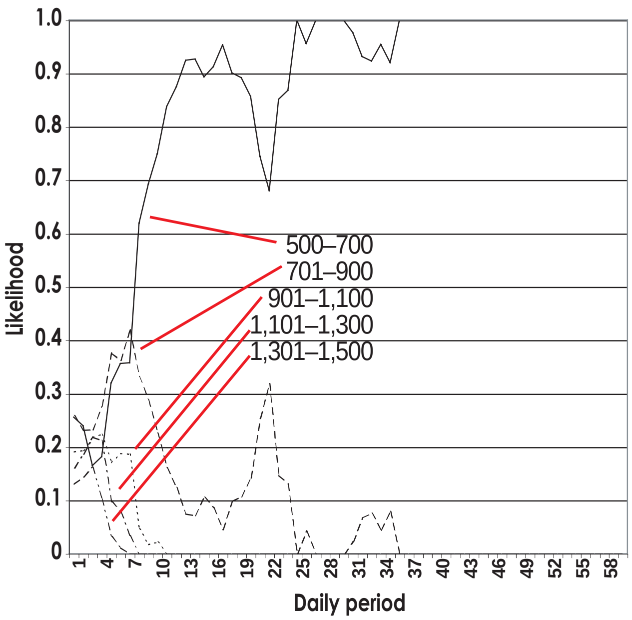
How Real-time Control Might Work in a Theme Park

The principles explained in the accompanying article were developed in restaurants, hotels, and theme parks. As I stated in the main text, a real-time control system (RTCS) is well-adapted to any service establishment that has high customer counts, including theme parks. In the case of a theme park, the RTCS would receive customer-arrival data from the park gates. Based on the day's weather, the RTCS would continually update its prediction about the business volume to be experienced throughout the day. The RTCS would also be fed real-time information from the payroll system—tracking which employees are late, or who have called in sick, for example. Finally the RTCS would receive real-time information from all point-of-sale systems within the park and from other data-tracking devices, such as queue-length monitors.

Using the current and predicted business volumes, the RTCS would serve as a management-decision aid: reoptimizing the labor schedule for the remainder of the day, recommending when to call in extra employees to work, when to send employees home, when and which employees to switch between positions to maximize the benefit to the organization, and when to send or recall employees from breaks. With complex hospitality service systems, like theme parks, RTCSs are the last, and presently largely uncharted, frontier of good labor management.

Exhibit 20

Business-volume likelihood, example 1



This graph displays the probabilities shown in the table in Exhibit 19. The level-1 line, for instance, comes from plotting the 25.56 percent of period 1, the 23.99 percent of period 2, and so on.

volume, 99 had customer counts that equaled or exceeded 45 customers *in period 10* and had 45 or fewer customers by the end of period 9. Level 2 had 30 comparative realizations, Level 3 had three such realizations, and Levels 4 and 5 saw no such comparable realizations. Thus, of the 132 comparable realizations, 75 percent resulted in a final daily customer count falling in level 1, just under 23 percent resulted in a final daily customer count falling in level 2, and slightly over 2 percent resulted in a final daily customer count falling in level 3. As of period 10, then, the manager has a strong indication that the final daily customer count will fall *below* level 3. By period 13 one can predict with a likelihood of over 90 percent that the day, *as a whole*, will experience level-1 demand.

The graph in Exhibit 20 displays the probabilities shown in Exhibit 19. The level-1 line in Exhibit 20, for instance, comes from plotting the 25.56 percent of period 1, the 23.99 percent of period 2, and so on. Showing the volume likelihoods graphically makes the day's demand easier to diagnose, or predict. Exhibits 21 and 22 illustrate other examples of how one might track such probabilities period by period throughout an operating day. Exhibit 21 shows that, as of period 20, there is greater than a 90-percent likelihood that the day, as a whole, will fall in business-volume-level 2. The customer-demand data illustrated in Exhibit 22 give a strong early indication (by period 15) that the day will hit level 3 or level 4. However, it is not until period 50 that the indication becomes clear that the day's demand will end up in level 4.

Exhibits 20, 21, and 22 raise the question of why one can predict the day-end business volumes earlier on some days than on others. The answer lies in the fact that initial customer counts may fall near the break points between categories. Turned around, the point is that the closer the final customer count is to the breakpoints between categories, the longer it takes to predict the final daily demand. Thus, if the customer counts are right on the cusp of two categories, the manager might not be able to establish the final daily demand until the end of the day. However, choosing one or the other of two adjacent categories is not the point of this process, so much as getting an early indication that the customer count will fall in one or the other of two adjacent volume categories. As in the case of Exhibit 22, one would have a consistent early signal that the day will likely be in one or the other of two adjacent volumes. In this case, the manager still can make the necessary real-time schedule adjustments, even if the certainty of the outcome isn't great.

The converse is also true: the stronger the indications are that final demand will fall in the middle of a category, the earlier in the day one can predict that day's business volume. Similarly, extreme volumes (i.e., level 1 or level 5) will be easier to predict than volumes falling in the mid-range.

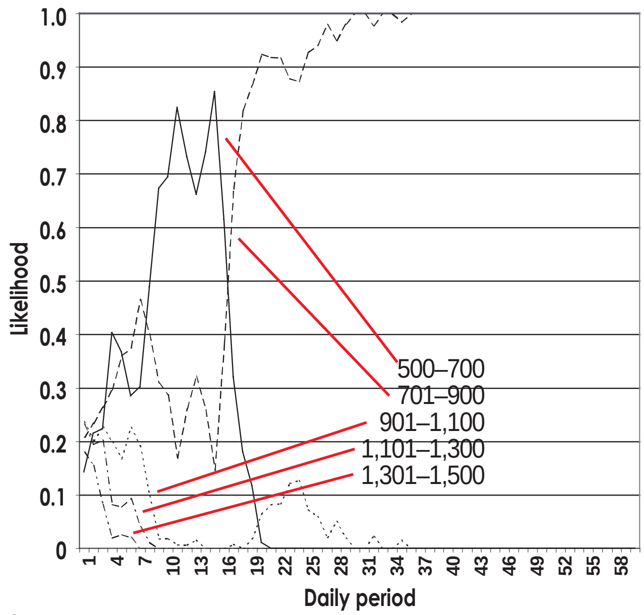
Real-time-control Actions

To develop the historical baseline data needed for this procedure, a manager should periodically perform steps 1 through 4. In contrast, step 5 should be performed hourly or even more frequently, because it is the monitoring step than allows one to predict day-end business volumes. In turn, predicting total daily business early in the day allows a manager to take appropriate long-lived actions to adjust employee schedules. Earlier in this report I explained how uncertain demand (and an easier ability to send employees home than to call them in) causes managers to inflate staffing levels. Thus, even with an expectation of average (level-3) demand, a manager probably would develop a schedule based on a level-4 volume just to be sure that all customers will receive appropriate service. Then, if the manager gets a strong early indication that demand will fall into a lower volume category, he can take appropriate long-lived actions (e.g., asking for volunteers to go home without pay). If, by contrast, he has set a level-4 schedule and then gets a strong signal that demand will hit level 5, he would want to take long-lived actions like extending employees' shifts, offering overtime, and perhaps calling additional employees in to work. He might even consider reoptimizing the day's labor schedule based on the new demand information.

On the other hand, if the operation is experiencing a real-time capacity-demand imbalance, but the indication is

Exhibit 21

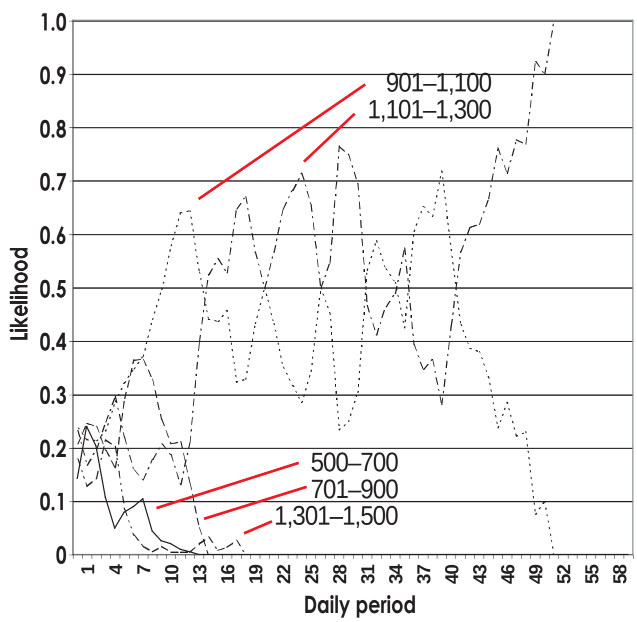
Business-volume likelihood, example 2



Showing the volume likelihoods graphically makes the day's demand easier to predict. In this example, as of period 20, there is greater than a 90-percent likelihood that the day, as a whole, will fall in business-volume-level 2 (701–900 total customers).

Exhibit 22

Business-volume likelihood, example 3



The customer-demand data illustrated here give a strong early indication (by period 15) that the day will hit level 3 or level 4 (901 to 1,300 total customers). However, it is not until period 50 that the indication becomes clear that the day's demand will end up in level 4 (1,101 to 1,300 total customers).

Exhibit 23

A comparison of staffing allocations under no cross utilization and cross-utilization scenarios

	Base Case—no cross training		**Scenario 1:** 20% wage and benefit premium for cross-trained employees				**Scenario 2:** 10% wage and benefit premium for cross-trained employees			
COV*	Staffing level**	Hourly cost	Staffing level	Hourly cost	Hourly savings	Percentage savings	Staffing level	Hourly cost	Hourly savings	Percentage savings
0.25	13/13/13/0	$403.81	9/9/9/6	$371.54	$32.27	7.99	9/9/9/6	$365.54	$38.27	9.48
0.20	12/12/12/0	378.51	10/10/10/3	360.41	$18.10	4.78	9/9/9/5	355.41	$23.10	6.10
0.15	11/11/11/0	357.59	10/10/10/2	349.77	$7.82	2.19	9/9/9/4	346.69	$10.90	3.05
0.10	10/10/10/0	347.62	10/10/10/1	341.26	$6.36	1.83	10/10/10/1	340.26	$7.36	2.12
0.05	10/10/10/0	335.71	10/10/10/0	335.71	0	0	10/10/10/0	335.71	0	0
0.00	10/10/10/0	333.12	10/10/10/0	333.12	0	0	10/10/10/0	333.12	0	0

*COV= Coefficient of variation of the forecast error.

**Best staffing levels are indicated are follows: position 1/ position 2/ position 3/ cross-trained employees.

not clear as to what volume the day will see, the only valid real-time-control actions are short-lived (e.g., reassigning employees). Taking long-lived actions runs the risk of invoking further (and otherwise unnecessary) actions later in the day.

The Value of Cross-Training

Even a relatively solid forecast contains the possibility of error, which is why managers usually err on the safe side and overstaff their operations. One way of reducing the effect of forecast uncertainty is by employing cross-trained workers. By having a cadre of cross-trained employees a manager gains scheduling flexibility, because she can deploy her cross-trained workers where they are most needed. Instead of counting the number of bussers and the number of runners, for instance, the manager could cross-train people for multiple jobs (including table servers) and set the schedule according to an estimate of the total help needed on the floor.

As an illustration of the value of cross-trained employees, consider the following example. Say that an operation has three different positions. Each of the positions would ideally be staffed by a complement of 10 employees, if the demand forecast were perfectly accurate. Exhibit 23 shows that with that perfect demand forecast the hourly cost of the system—both labor costs and the cost of customer waiting—would be $335.71.[32] As the uncertainty in the demand forecast increases (as measured by the coefficient of variation of forecast error), the ideal staffing level increases to 13 employees per position and the total hourly cost rises to $403.81.

Exhibit 23 also shows the effect on hourly costs and staffing decisions when a manager can draw from a pool of cross-trained employees. Exhibit 23 considers scenarios where the cross-trained employees receive pay premiums

[32] For a discussion of how to calculate the cost of customers' waiting time, see: Thompson, Part 2, p. 32.

of either 20 percent or 10 percent compared to other employees. Assuming a 20-percent wage-and-benefit premium for the cross-trained employees, the ideal allocation of employees under the highest level of forecast inaccuracy would be to assign nine employees to each of the three positions and have six cross-trained employees who would be assigned in real time to the positions so as to balance the workload. This labor allocation would require 33 employees in total and cost $371.54 per hour, representing a 15-percent reduction in the number of employees and an 8-percent cost saving compared to staffing the positions with dedicated employees.

A close examination of the results in Exhibit 23 reveals several patterns. First, without cross-trained employees, higher forecast inaccuracy leads to **(a)** higher staffing levels and **(b)** higher hourly costs. Second, using cross-trained employees, higher forecast inaccuracy yields **(a)** a larger number of cross-trained employees and lower numbers of position-specific employees and **(b)** greater savings from cross-trained employees. Finally, when the cross-trained employees are relatively less expensive (than regular workers), **(a)** more cross-trained and fewer position-specific employees are warranted and **(b)** larger savings accrue from having cross-trained employees. Indeed, under the highest level of forecast inaccuracy, having a pool of cross-trained employees reduced employee needs by 15 percent and reduced costs by over 9 percent. Although it is not shown in Exhibit 23, the benefit of cross-trained employees is also greater when the employees can be shared across more than two jobs.

The availability of a cross-trained labor pool should be incorporated during the development of a labor schedule (discussed in the previous section of the report). However, the actual deployment of the cross-trained labor would occur in real time, when employees would be assigned to the positions most useful to the hospitality firm.

Conclusion

This report has focused on the task of workforce scheduling in hospitality businesses. My hope is that it has given you a better understanding of the complexity of the task, but also that you've come away with some ideas about how to better manage the scheduling process in your firm. Based on what I've seen, hospitality firms have many opportunities to improve their workforce-scheduling process. ■

About the author

Gary M. Thompson, Ph.D., is a professor of operations management at the Cornell University School of Hotel Administration (gmt1@cornell.edu). His chief area of research is labor scheduling in hospitality operations.

Cornell Hotel and Restaurant Administration Quarterly

Hospitality's Premier Management Publication!

Whether you are a CEO, an operations manager, an executive trainee, or an undergraduate, there is something for you in every informative issue of

CQ

Cornell Hotel and Restaurant Administration Quarterly!

Here are some of the vital topics covered in recent issues:

- How Best Practices Evolve—and Improve Your Operation
- How to Drive Customers' Loyalty by Gauging Satisfaction
- How to Benchmark Your Most Effective GMs
- How to Improve Revenue Management through Process Control
- The Danger of Personal Bias in Making Revenue Forecasts

"As both a contributor and reader, I have found the Cornell Quarterly to be a source of powerful insights that have great practical value."

Peter Yesawich, Ph.D.
President and CEO, Yesawich, Pepperdine, Brown & Russell

"The Cornell Quarterly is a leading source of new lodging and food-service topics and ideas that are both conceptual and practical for planning and operations."

Bjorn Hanson, Ph.D.
Global Industry Managing Partner, PricewaterhouseCoopers

Keep ahead of the world's fastest-moving industry with

Hospitality's Premier Publication!

Subscribe to the **Cornell Quarterly** today!

Contact:
SAGE Publications
2455 Teller Road, Thousand Oaks, CA 91320 U.S.A.
Telephone: (800-818-7243 (U.S.) • (805)-499-9774 (Outside of the U.S.)
FAX: (800-583-2665 (U.S.) • (805)-499-0871 (Outside of the U.S.)
Email: journals@sagepub.com or visit www.sagepub.com