



ENHANCING EMPLOYEE AND ORGANIZATIONAL PERFORMANCE THROUGH COACHING BASED ON MYSTERY SHOPPER FEEDBACK: A QUASI-EXPERIMENTAL STUDY

**GARY P. LATHAM, ROBERT C. FORD,
AND DANNY TZABBAR**

Based on reinforcement theory, a quasi-experimental design was used to evaluate the effect of (a) feedback obtained from (b) a relatively neutral third party (namely, mystery shoppers) that was obtained on a (c) variable interval schedule for managers to use to (d) coach their employees. An interrupted time-series design showed that both employee and organizational performance increased as a result of this intervention. Performance dropped when this intervention was cut back and, subsequently, discontinued. These results were replicated in two additional restaurants. © 2012 Wiley Periodicals, Inc.

Keywords: employee coaching, mystery shopping, variable interval schedule, performance feedback

Introduction

The service industry is a major driver of Western economies. In the United States alone, it represents over 75 percent of the GDP (Ford & Bowen, 2008). Thus, it follows that research directed at increasing an understanding of drivers of employee and organizational performance in this sector is of growing concern to human resource management scholars and practitioners (e.g., Bowen & Ford, 2002;

Chesbrough & Spohrer, 2006; Liao, 2007; Pugh, 2001; Schneider, Ehrhart, Mayer, & Saltz, 2005). The purpose of the present study was to contribute to this knowledge base through a quasi-experimental investigation.

Specifically, this study examined the application of a technique adapted from marketing, namely, mystery shopping, to overcome a problem that confronts most managers—providing employees timely systematic

Correspondence to: Robert C. Ford, College of Business Administration, 400 Central Florida Blvd., Orlando, FL 32816, Phone: 407.823.5088, E-mail: rford@bus.ucf.edu

Human Resource Management, March–April 2012, Vol. 51, No. 2. Pp. 213–230

© 2012 Wiley Periodicals, Inc.

Published online in Wiley Online Library (wileyonlinelibrary.com).

DOI:10.1002/hrm.21467

performance feedback. Specifically, we looked at the effect of feedback from a third party unknown to both the employees and their supervisors—hence, the term *mystery*. Mystery shoppers assessed the performance of employees in three restaurants on a variable interval schedule. We wanted to see if feedback from this source could be used to enhance the coaching effectiveness of supervisors on the performance of frontline sales staff who in turn have a direct impact on their organization's performance. Prior

Extensive empirical research conducted in experimental and organizational psychology throughout the twentieth century has demonstrated that feedback is critical for learning (i.e., ability) and the choice to exert effort and persist in goal attainment (i.e., motivation).

to conducting this study, there was no evidence to suggest that employee performance and the performance of the restaurant where they worked would increase or decrease as a result of this intervention. Thus, we wanted to ascertain whether supervisory coaching of direct reports based on feedback provided on a seemingly random, albeit systematic schedule from a neutral third party affects employee and organizational performance positively.

Performance Management

Performance management in the service industry is typically problematic. Requiring a manager to systematically assess each employee is often not logistically possible. In the majority of instances, an employee cannot consistently be observed interacting with a customer who is being served. None-

theless, employees in this new millennium want relatively immediate performance feedback (DeStobbeleir & Ashford, in press). Empirical research shows that this request is justified. Extensive empirical research conducted in experimental and organizational psychology throughout the twentieth century has demonstrated that feedback is critical for learning (i.e., ability) and the choice to exert effort and persist in goal attainment (i.e., motivation) (e.g., Ammons, 1954; Annett, 1961; Ilgen,

Fisher, & Taylor, 1979; Locke & Latham, 2002).

Feedback/knowledge of results that is delayed is far less effective than relatively immediate feedback for facilitating learning and maintaining effort and persistence. Because performance appraisals are typically done on a fixed interval basis—that is, annually—the feedback an employee may need to act upon to improve performance is not given in a timely fashion. Thus, it should not be surprising that there is little or no evidence for the beneficial effect of performance appraisals on job performance (Latham & Mann, 2006).

Ongoing coaching, in addition to performance appraisals, is viewed by many practitioners (e.g., Hutcheson, 1996; Luthans & Peterson, 2003; Waldroop & Butler, 1996) as a primary way to develop and motivate employees. This is because coaching overcomes the problem of timeliness. It involves giving feedback on an ongoing basis. Based on this feedback, specific goals are set, the relationship between what the person is doing and the outcomes the person can expect is clarified, good performance is praised, and the individual is inspired to take action that will result in an improvement in job performance (Hall, Otazo, & Hollenbeck, 1999; Heslin, Latham, & VandeWalle, 2005; Heslin, VandeWalle, & Latham, 2006).

Despite the fact that coaching has become a commonplace prescription in the popular press for managing the performance of employees, there is a paucity of empirical research on the effectiveness of this process. Exceptions to this statement include a study by Sue-Chan and Latham (2004). They examined the relative effectiveness of outsider, peer, and self coaching on the team-playing behavior of MBA students in their respective study groups. Smither, London, Flautt, Vargas, and Kucine (2003) examined the importance of goal setting and providing feedback in relation to goal pursuit. Heslin et al. (2006) examined the implicit person theory held by managers regarding the malleability of behavior. Yet no study to date has examined ways of overcoming the inability of managers, especially in the service industry,

to find the time to provide ongoing feedback systematically to their employees. There is a critical need to do so.

Komaki (1994) found that in practice supervisors spend less than 1 percent of their time observing their subordinates. This is because, as Ashford and Northcraft (2003) argued, managerial work has been organized in ways that have increased managers' span of control. A consequence of a wide span of control is that accurate and timely supervisory appraisals of employees are an increasingly difficult undertaking. As a result, many employees often view appraisals of their performance as biased, and empirical research supports this perception (Lance, 1994; Lefkowitz, 2000; Strauss, Barrick, & Connerley, 2001). There is a theory and a technique, however, that suggest a solution to the issues of timeliness and bias.

Human Resource Management Theories

At least three human resource management theories of organizational behavior stress the necessity of providing employees feedback on their performance. Feedback is a moderator variable in goal setting theory (Latham & Locke, 2007; Locke & Latham, 2002) because feedback is necessary for effectively guiding goal pursuit. Social cognitive theory (Bandura, 2001) emphasizes the importance of feedback for increasing self-efficacy for goal attainment. The theory also explains the importance of feedback for enabling employees to see the relationship between what they are doing and the outcome they can expect (e.g., goal attainment). Neither theory, however, specifies the frequency with which feedback should be given. The answer is suggested in Skinner's (1974) theory of reinforcement.

Although the philosophy underlying reinforcement theory—namely, behaviorism (Watson, 1924)—has been discredited for failing to acknowledge cognition as a mediating variable (Bandura, 2001; Latham, 2007; Locke & Latham, 1990), this failure does not negate the effectiveness of this theoretical framework for suggesting ways of managing performance (Dunnette, 1976). Voluminous experiments

show that when learning a response, a continuous schedule of reinforcement results in higher performance than a variable schedule (Ferster & Skinner, 1957; Latham & Dossett, 1978). Once the response is learned, and reinforcement is subsequently administered on a fixed interval such as once every minute, the responses increase rapidly only as the end of the predetermined time period approaches (e.g., reinforcement is only available at the end of 60 seconds), and drops rapidly to near zero immediately following the reinforcement. But, on a variable interval (VI) schedule, responses are emitted at a high steady rate as it becomes increasingly difficult to discern when the seemingly random time interval will terminate and reinforcement will be available. No study to the authors' knowledge has investigated the effectiveness of a VI schedule on the job performance of employees and the organization that employs them.

Providing feedback on a VI schedule does not eliminate the problem of managers finding the time to coach all their employees on a systematic basis. It only suggests that there may not be a necessity for coaching all employees on a continuous basis.

Mystery Shoppers

For upper-level managers, the coach is often a professional consultant who is not employed by the organization (Hall et al., 1999). This "neutral" party often "shadows" the manager on the job. The hiring of outsiders, however, is prohibitively expensive for coaching employees in lower-level jobs. Luminaries in the practitioner world, such as Jack Welch, the former CEO of the General Electric Company, have argued that coaching is a core competency of leadership. Effective leaders, he said, are those who find the time to "grow" their people (Welch, 2001). The critical phrase is "to find the time."

No study to date has examined ways of overcoming the inability of managers, especially in the service industry, to find the time to provide ongoing feedback systematically to their employees. There is a critical need to do so.

Changing the word *appraiser* to *coach* does not resolve the issue of enabling managers to find the time to systematically observe the performance of each of their direct reports. While a baseball coach is standing in the dugout observing the ongoing performance of the players on the field, and a hockey coach is standing behind the bench watching players skate, the typical manager/coach in industry and government is likely to be away from employees performing myriad job duties (e.g., strategic planning, coordinating implementation with other department heads). Thus, the decision was made by senior management of a restaurant

A solution to the problem of managers not being able to systematically observe their employees might be to employ a third party to do so—namely, a mystery shopper.

chain to employ mystery shoppers to provide relatively immediate feedback for coaching employees on a timely basis.

Mystery shoppers are typically hired by organizations in the service industry to assess employee–customer interactions (Wilson, 1998). The dependent variable is usually customer satisfaction. Mystery shoppers are generally given a standardized form, based on a job analysis, to complete regarding their observations of an employee. This allows assessments across time and units (Beck & Miao, 2003; Finn, 2001; VanderWiele, Hesselink, & Van Iwaarden, 2005).

The effectiveness of mystery shopping is due in part to the fact that no one in the organization knows the identity of a mystery shopper (Lewis & Chambers, 2000). Thus, the employees do not know when they are being assessed. Nor is the time period for the assessment known because it occurs on a predetermined VI schedule. Hence, a solution to the problem of managers not being able to systematically observe their employees might be to employ a third party to do so—namely, a mystery shopper. Using mystery shopper feedback on an employee's performance for coaching purposes has never before been investigated.

The overarching hypothesis of the present investigation was that mystery shopping

is an effective method for resolving difficulties inherent in coaching, not to mention the traditional performance appraisal. This is because mystery shopping provides relatively immediate feedback systematically on a variable interval basis, the feedback is relatively objective in that it is based on observations of behavior, and it is directly tied to important organizational outcomes identified through job analysis. Moreover, the feedback given to an employee by a supervisor comes from an external source that is relatively unbiased, as the mystery shopper has never before interacted with the employee.

In summary, this was the first study to investigate the effectiveness of supervisory coaching based on (a) feedback from a third party (b) provided on a variable interval schedule to employees. Two hypotheses were tested: (1) supervisory coaching, using feedback from a relatively neutral third party that is provided on a variable interval schedule, increases employee performance and (2) this coaching process also increases an organization's performance.

Method

Context

Three restaurants in the same restaurant chain, located in the same metropolitan area, participated in this quasi-field experiment. The restaurants can be characterized as “casual dining” as opposed to “quick serve” or “fine dining.” The seating capacity ranges from 225 to 275 seats. The restaurants serve both lunch and dinner. Each restaurant serves customers 11 hours a day, seven days a week. The sales revenue of each restaurant falls in the range of \$3 to \$4 million annually.

Each of the three restaurants in this study employed approximately 30 servers.

Servers are a critical factor affecting the profitability of a restaurant (Andaleeb & Conway, 2006; Gupta, McLaughlin, & Gomez, 2007). This is because they greet customers, suggest/take orders, check the presentation of food for both appearance and accuracy prior to serving it, and present and process the bill in a (un)timely manner. The customer–server

interaction affects “repeat business” (Carrillat, Jaramillo, & Mulki, 2009; Ford, Wilderom, & Caparella, 2008; Paul, Hennig-Thurau, Grewler, Gwinner, & Wiertz, 2009).

Prior to conducting this study, the restaurant chain had trained the servers on the importance of, and ways to engage in, the behaviors critical to job performance/customer satisfaction. These behaviors had been identified through the company’s job analysis of performance. Thus, the restaurants’ servers had the knowledge, skills, and ability to perform their jobs.

Mystery Shopping Procedure

The mystery shoppers were drawn from a national database of approximately 300,000 households developed and maintained by the mystery shopping firm hired by the restaurant chain. The shoppers matched the customer demographics of the restaurant chain’s target market—namely, 25–55 years of age, a minimum of some college education, and a minimum income of \$50,000. They are people who typically eat in a similar restaurant at least three times a month but who are not employed as restaurant consultants or critics. A shopper receives no training, as the employing company wants observations from a “typical customer.” The shoppers in this study were reimbursed up to \$35 of their dining bill.

Each mystery shopper received an e-mail from the shopping consulting firm as to the specific date and hour to eat at one of the three restaurants. The e-mail also specified the categories of food the shopper was to eat (e.g., must order an entrée and at least one appetizer). No shopper visited the same restaurant more than once. Each shopper was accompanied by one to three friends or relatives. No server and no manager in any of those restaurants knew the hour when a shopper would appear or the person’s identity. Hence, a shopper’s presence in a restaurant was on a variable interval-type schedule. On average, each employee could expect to serve a mystery shopper once a month, as a shopper visited a restaurant daily.

Upon leaving the restaurant, the mystery shopper completed a 110-item questionnaire designed to evaluate the behaviors of the server who was identified by a name tag worn on the server’s uniform. The questions assessed the desired behaviors of a server that had been identified through the job analysis. These questions focused on three major areas: (1) *time and timing* (e.g., “How long before you were greeted by your server?” “Did your server get back to you within a reasonable time period to check on your meal?”); (2) *professionalism* (e.g., “Was the server attentive to your needs by anticipating things before you had to ask?” “Was the server polite and careful not to interrupt your conversation unnecessarily?” “Was the table kept clean and cleared of any unnecessary plates, glassware, and trash throughout the meal?”); and (3) *knowledge* (e.g., “Was your server knowledgeable of the menu items and wine list without overselling it?” “Was dessert offered after the meal with a specific suggestion?”). Each of the 110 items allowed a shopper to assess a server’s behavior (mostly on a two-point scale [i.e., yes/no]), with several items also allowing a response of N/A and some items with different anchors, such as for “time to greet,” for which the possible responses were “right away,” “too slowly,” or “N/A.”

There was a 100 percent return rate from the mystery shoppers. This is likely due to the fact that they would not have been reimbursed if they did not post their report within the time frame mentioned earlier, nor would they have been rehired by the shopper consulting firm.

The mystery shopper’s report on an individual server was posted on the respective restaurant’s secure online reporting site within 12–18 hours of the visit. A manager provided the server with feedback from the shopper prior to the start of the next day’s shift, praised effective behaviors, and set behavioral

Upon leaving the restaurant, the mystery shopper completed a 110-item questionnaire designed to evaluate the behaviors of the server who was identified by a name tag worn on the server’s uniform.

goals for improvement (e.g., suggest dessert to each customer, greet customer within three minutes of seating). Hence, the feedback to and goal setting for an employee was relatively timely and objective.

To ensure the manager provided the mystery shopper feedback from a motivational/developmental vantage point, the manager was trained by the organization's training department to (1) acknowledge/praise a server for the effective behaviors that were observed by a shopper, (2) ask the server for ideas on how to improve performance in areas that were identified by the mystery shopper as performed poorly, (3) set a behavioral goal based on the feedback, (4) discuss ways of dealing with difficult customers or situations, and (5) summarize with a server action steps for obtaining a score of 90 percent or higher from future mystery shoppers. In short, the emphasis of a coaching session was motivational/developmental rather than critical. No incentive other than feedback from a mystery shopper and the setting of behavioral goals was administered to the servers.

At a later point in the study, the number of mystery shoppers was reduced by two-thirds—namely, to 10 visits a month to each restaurant instead of 30. This was done because senior management believed that the performance of the three restaurants involved in this study had reached a high level.

Research Design and Data Collection

The data were analyzed with an interrupted time-series quasi-experimental design (Campbell & Stanley, 1972). The design involved multiple time-interval pretest measures (1) when no shopper feedback was provided to any employee, (2) the intervention (coaching based on a mystery shopper's feedback), (3) a two-thirds decrease in mystery shopping feedback, and (4) a posttest performance measure where no shopper feedback was provided to the servers. Data on two performance measures were obtained from three different restaurants of the same restaurant chain for 40 months. Mystery shopping

was initiated in the three restaurants at different points in time. Thus, there is variance in the data-collection periods for each stage in the study, as described below.

Precoaching Time Period

Consistent with the recommendations of Campbell and Stanley (1972), mystery shopping data on server performance were collected prior to the coaching based on mystery shopping feedback. Specifically, 16 months of premeasure performance of customer count was collected for restaurants 1 and 2, and 17 months of premeasure performance of customer count was collected for restaurant 3.

Coaching Intervention

The duration of mystery shopping was approximately a year (12 months for restaurants 1 and 2, and 11 months for restaurant 3). In month 28, the frequency of mystery shopping feedback to servers was reduced by two-thirds (i.e., from 30 to 10 mystery shopping visits a month) in all three restaurants. The reduction in coaching based on mystery shopping feedback lasted for five months (i.e., until month 33).

Postcoaching Period

Seven additional months of data were collected (i.e., until month 40) on each restaurant's customer count. No mystery shopper feedback was available for the servers during this time period.

Customer count was recorded by the restaurant for 16 months prior to the introduction of the mystery shopper feedback interval and 8 months after it had been terminated. The data parallel those obtained for server performance.

Employee/Server Performance

The mystery shopper's behavioral observation scores of a given server were the sole rating of that employee's performance. The survey was extremely detailed, with 110 separate questions assessing the entire customer

experience. The percentages were calculated by dividing a mystery shopper’s actual scoring of a server’s performance by the maximum possible score (excludes NAs). The percentages for each time period were the sum of all daily mystery shopper scores divided by the number of days in the period. Thus, possible scores ranged from 0 to 100 percent.

The data on server performance were collected for 21 months—namely, from month 13 through month 30. The months in which data were collected prior to the mystery shopping intervention (i.e., until month 17 for restaurants 1 and 2, and until month 18 for restaurant 3) were used as baseline data. That is, no feedback from a mystery shopper was provided to any employee during this pre-intervention time period. This baseline was compared with the data collected during the subsequent time period when feedback from mystery shoppers was provided by the restaurant manager to a server (i.e., until month 28). Server performance during the baseline and the intervention time periods were also compared with the following five-month time period when mystery shopping feedback to servers was reduced by one-third (i.e., month 29 to month 33). A limitation of these data is that server performance was not assessed by the organization after coaching based on mystery shopping feedback was terminated. Only each restaurant’s customer count was assessed.

Restaurant Performance

To ensure confidentiality, senior management provided us with a percentage change of customer count instead of the actual head count. The baseline for the percent change was based on the average count of customers for the year prior to the start of this study. The baseline varied for each of the three restaurants. The percentage change reflects an increase, decrease, or no change relative to the baseline. Specifically, each month of the study was compared with the baseline established in the year prior to the intervention. For example, if the average headcount baseline in year $t-1$ was 10,000 customers, months 1, 2, 3 . . . 40 of the study were compared with

this same baseline number. In that case, a 10 percent increase in month 12 means that the restaurant had 11,000 customers that month. Similarly, a 10 percent decrease reported in month 4 reflects a count of 9,000 customers for that month.

Data Analysis

Bergh (1993) argued that the analysis of an interrupted time-series design relies on differential responses over time to reveal if there is a treatment effect. With a large number of observations across time intervals, the sources of variation can be determined with a high degree of reliability. A “panel” of data is a set of observations with both “time-series” and “cross-section” dimensions. A “time series” arises from observations on the same unit (e.g., restaurant) at different points in time (e.g., months). A “cross-section” arises from contemporaneous observations of the dependent variable(s) on different units. In this study, there were two time series—namely, 30 monthly observations of servers’ performance and 40 monthly observations of restaurant customer count. Accordingly, we conducted a fixed effect generalized least square test to assess the effect of coaching, based on mystery shopping feedback, on employee and restaurant performance (Hsiao, 1986; McDowall, McCleary, Meidinger, & Hay, 1980).

A limitation of time-series analysis is that it often suffers from temporal autocorrelation in that successive observations are not statistically independent. This results in autocorrelation of the error terms that in turn increases the risk of overestimating the statistical significance of the intervention (i.e., coaching). A technique that takes this structure of the error terms into account is the ARIMA method developed by Box and Jenkins (McDowall et al., 1980). To supplement our generalized least squares analysis (GLS) model, we used interrupted ARIMA time-series analyses. The ARIMA model allowed a comparison between

To ensure confidentiality, senior management provided us with a percentage change of customer count instead of the actual head count.

the pre-, during-, and postcoaching time periods. The data were analyzed using STATA, invoking the robustness function, and using a one-month lag for each dependent variable.

Results

The percentage changes in server performance and customer count over the duration of the study are presented in Figures 1 and 2, respectively. The two figures indicate that, on average, the three restaurants had higher server performance and a higher customer count in the period when the frequency of coaching, based on mystery shopper feedback, was provided daily. The two figures also show that server performance and customer count were lower in the case of the pre- and postcoaching time periods when no feedback from a mystery shopper was made available to the servers.

Consistent with both Sawyer, Latham, Pritchard, and Bennett (1999) and Bergh's (1993) recommendations, the analysis of these data followed a two-step process. First, tests for the assumptions underlying the time-series model were conducted. Second, the time-series panel analysis was performed to test the hypothesized relationships.

A test for multicollinearity showed that the variance inflation factor for all variables

was quite low (max = 0.25), well below the acceptable level of 10. Thus, multicollinearity does not appear to have affected parameter estimates for the regression parameters. Our initial analysis confirmed the importance of controlling for time-series and cross-section effects in the data. When cross-section and time-series components of the error were not controlled, (1) the hypothesis of first-order autocorrelation could not be rejected (as indicated by a Durbin-Watson statistic of 2.03) and (2) the hypothesis of homoscedasticity could be rejected (as indicated by an F statistic of 3.23 for White's test). Because autocorrelation and heteroscedasticity were evident in our data, our construction of a model with a time-series component in the error structure was necessary. Accordingly, the Fuller-Battese GLS approach with time-series and cross-section effects was used to adjust for autocorrelation in the time series and for heteroscedasticity in the cross-section.

When the analysis was appropriately placed into a panel framework (i.e., with an error structure that allows estimation of the time-series and cross-section effects), and is estimated with generalized least squares, the model conformed with Bergh's (1993) guidelines for the appropriate treatment of data with both time-series and cross-sectional components. Thus, a generalized least squares

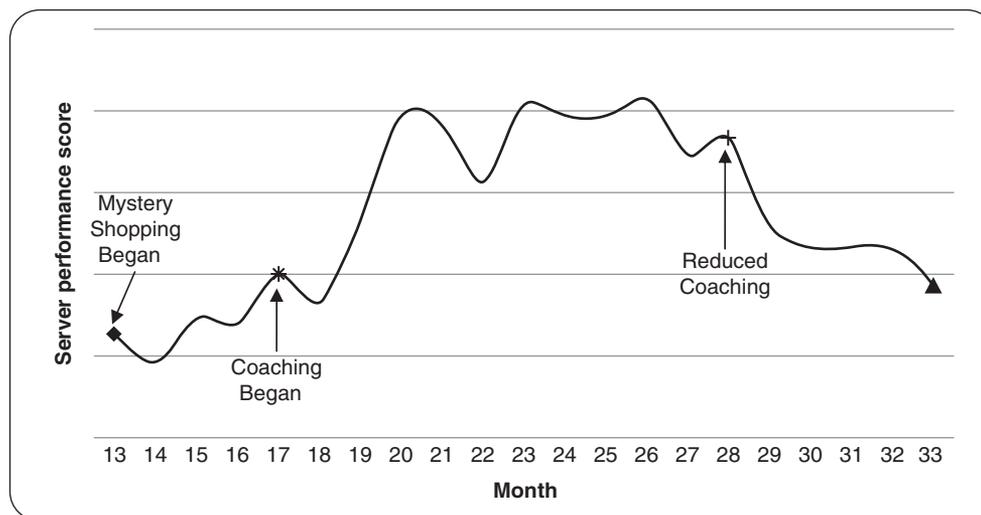


FIGURE 1. Average Server Performance Across Three Restaurants

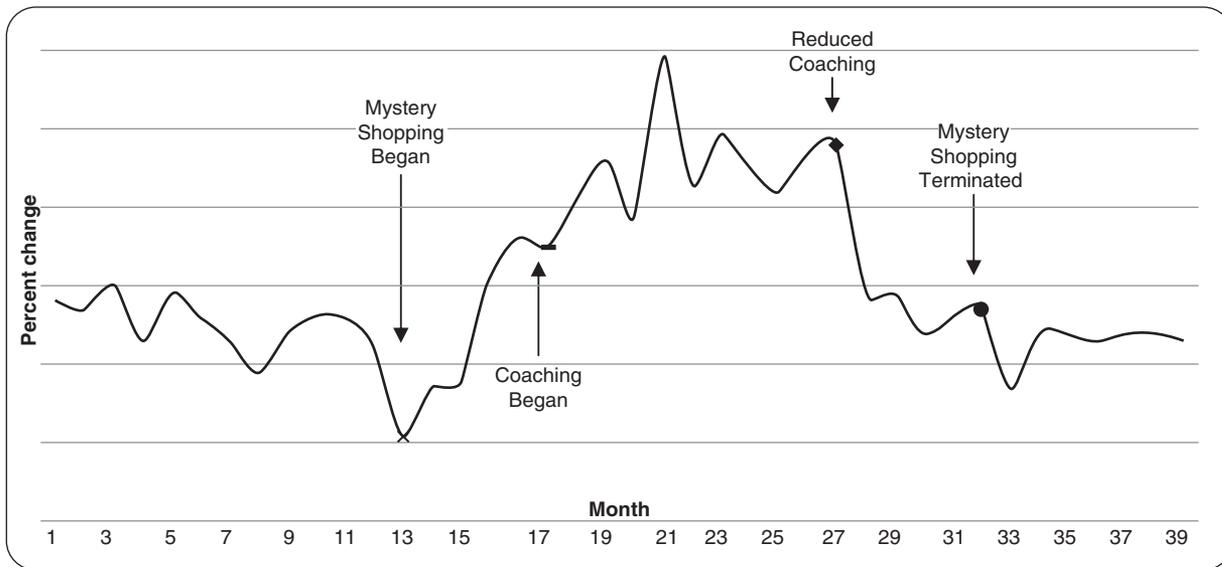


FIGURE 2. Percentage Change in Average Customer Count Across Three Restaurants

estimation was conducted using a Fuller-Battese Variance Components Model. As shown in Table I, the panel framework, estimated with the Fuller-Battese procedure, used the time-series effects in the data to obtain unbiased, consistent, efficient estimates of the coefficients. The GLS estimate of variance associated with the time dimension (i.e., the months) was 3.91 and 3.69 for server performance and customer count, respectively. The variance associated with cross-section (indicators and restaurants) was 17.79 and 17.89. The variance associated with error was 83.03 and 83.63. Thus, the variance was dominated by the residual error. The cross-sectional effects across indicators and restaurants, while small, are much larger than the time-series effects. This indicates little variance due to the time dimension beyond that accounted for by month-to-month change in server performance. The lagged server performance and customer count scores were included in our model.

Table II shows the results of the GLS. These results provide strong support for the two hypotheses. Feedback from mystery shoppers was positively and significantly related to both server performance and a restaurant's customer count. The model explains

40% of the variance in server performance (adjusted $R^2 = .40$, $p < .05$). This finding is statistically significant ($F = 15.10$, $p < .05$). Similarly, the model explains 48 percent of the variance in the month-to-month changes in customer count (adjusted $R^2 = .48$, $p < .05$). This too is statistically significant ($F = 18.44$, $p < .05$).

Tables IIIA and IIIB and IVA and IVB present the results of the ARIMA model on customer count and server performance, respectively. Specifically, in Table IIIA, we explore the relationship between coaching and customer count. To make sure our results were due to coaching, and not the use of mystery shopping, we first entered the dummy variable associated with the period of mystery shopping in Model 1, and then added the dummy variable associated with coaching in Model 2. As shown in Model 2 in Table IIIA, only coaching was positively and significantly related to the increase in customer count in all three restaurants, providing additional credence to the initial findings reported earlier.

To evaluate the degree to which the coaching intervention is significantly different from both the pre- and postcoaching period, we conducted an additional analysis reported in Table IIIB. In Model 1, we compared

TABLE I Variance Component Estimates From the GLS Fuller-Battese Method

	Server Performance Model	Customer Count Model
Root mean squared error	8.37	9.27
Variance component for cross-section	17.79	17.89
Variance component for time series	3.91	3.69
Variance component for error	83.03	83.63

TABLE II GLS Fixed-Effect Model for the Relationship Among Coaching, Server Performance, and Restaurant Customer Count

	Server Performance	Customer Count
<i>Coaching</i>	1.68* (.65)	5.23* (2.12)
<i>Constant</i>	-15.10	-18.44*
<i>R² overall</i>	.40	.48

Standard error in parentheses; * $p < .05$.
 $N = 120$.

TABLE IIIA ARIMA Model for the Relationship Between Coaching and Customer Count*

	Restaurant 1		Restaurant 2		Restaurant 3	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Mystery shopping	14.198* (3.07)	2.02 (3.84)	14.85* (3.11)	4.36 (3.49)	12.90* (3.72)	-.16 (5.12)
Coaching		21.31* (4.09)		18.36* (4.27)		22.86* (5.10)
<i>Constant</i>	-4.58*	-4.58*	-4.47*	-4.47*	-2.95*	-2.95*
<i>/sigma</i>	9.99*	6.45*	10.09*	7.64*	11.98*	8.74*
<i>Log likelihood</i>	-148.86*	-131.33*	-149.20*	-138.10*	-156.10*	-143.48*

* β coefficients are reported and standard errors in parentheses.

TABLE IIIB ARIMA Model Comparing Pre-, During-, and Postcoaching Effects on Customer Count*

	Restaurant 1		Restaurant 2		Restaurant 3	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Coaching	25.75* (2.02)		24.92* (3.02)		26.22* (3.10)	
Postcoaching	6.62* (2.39)		7.74* (2.34)		7.47* (3.44)	
<i>Constant</i>	-7.01*		-6.66*		-6.47*	
Precoaching	-26.07* (1.99)		-25.09* (2.88)		-29.03* (2.63)	
Postcoaching	-19.13* (2.59)		-17.17* (3.33)		-18.75* (2.58)	
<i>Constant</i>	18.75*		18.25*		19.75*	
<i>/sigma</i>	5.88*		7.13*		8.17*	
<i>Log likelihood</i>	-127.65*		-135.34*		-140.76*	

* β coefficients are reported and standard errors in parentheses.

the coaching and postcoaching periods with the pre-coaching period. As shown in Model 1, customer count is significantly higher in the coaching and postcoaching period relative to the pre-coaching period, and the effect of coaching on customer count is significantly stronger than the postcoaching period. In Model 2 in Table IIIB, we compared pre- and postcoaching periods with the coaching period. As shown, the customer count is significantly lower in the pre- and postcoaching time periods relative to the time period in which coaching employees took place.

Table IVA reports the ARIMA model exploring the relationship between coaching and server performance. As shown, coaching is associated with a significant increase in server performance in all three restaurants.

Table IVB compares the different periods. Model 1 shows that server performance is significantly higher during the coaching and postcoaching periods relative to the pre-coaching period. Server performance, however, is significantly lower in the pre- and postcoaching periods relative to the coaching period, as shown in Model 2.

Overall, Models 1 and 2 in both Tables IIIB and IVB indicate that customer count and server performance is lower during the post-coaching relative to the coaching period, and is significantly higher than the pre-coaching period. This suggests a relatively long-term effect of coaching on employee and organizational performance.

To evaluate the financial benefit of the mystery shopper program, we requested

TABLE IVA ARIMA Model for the Relationship Between Coaching and Server Performance*

	Restaurant 1	Restaurant 2	Restaurant 3
Coaching	13.42* (2.23)	10.69* (1.96)	8.17* (2.24)
<i>Constant</i>	79.01	82.22	86.67
<i>/sigma</i>	4.46*	3.95*	4.54*
<i>Log likelihood</i>	-61.18*	-58.67*	-61.59*

* β coefficients are reported and standard errors in parentheses.

TABLE IVB ARIMA Model Comparing Pre-, During-, and Postcoaching Effects on Server Performance*

	Restaurant 1	Restaurant 2	Restaurant 3
	Model 1	Model 1	Model 1
Coaching	20.08* (2.49)	15.58* (1.64)	14.5* (1.34)
Postcoaching	10.01* (2.89)	7.33* (2.44)	9.50* (2.24)
<i>Constant</i>	72.33*	77.33*	80.33
	Model 2	Model 2	Model 2
Precoaching	-18.12* (2.54)	-13.23* (1.87)	-19.3* (1.29)
Postcoaching	-10.08* (1.81)	-8.25* (2.07)	-5.05* (2.19)
<i>Constant</i>	92.42*	92.92*	94.83*
<i>/sigma</i>	3.21*	3.24*	3.47*
<i>Log likelihood</i>	-54.32*	-54.51*	-55.94*

* β coefficients are reported and standard errors in parentheses.

the data senior management actually used to evaluate the program. The vice president in charge of this program reported:

These data showed both direct and indirect benefits of the mystery shopper program. It yielded nearly \$2 in profit for every \$1 spent on the program. There were direct benefits resulting from the amount spent by the mystery shoppers themselves in excess of their reimbursement. The average amount spent in the restaurant by each mystery shopper was calculated at just over \$22 per person. Since each mystery shopper was required to bring at least one additional person (and many brought more than one additional person) to the restaurant, the revenue yield from each mystery shopper party actually averaged almost \$80 per visit. The designated shopper was reimbursed a total of \$35 per visit, and the mystery shopper firm was paid a \$15 management fee, making the total cost to the company of \$50 per visit. Thus, with an incremental gross margin of almost 45 percent, the cost of the shopper program was almost completely covered by the revenues generated by that visit.

The indirect benefits of this program were as follows. The mystery shoppers returned with a surprisingly high frequency of 2.4 times in the following 12 months—with no financial incentive. Additionally, they seldom returned alone, averaging more than 2.5 guests per party. These incremental visits were very profitable and more than offset all the costs of the program and yielded an additional incremental profit of \$44 for each mystery shopper visit! Effectively, the mystery shopper program became a superb method of introducing customers to the concept, which worked out to have an approximately two-for-one profit-to-cost ratio.

A second indirect benefit is one that is more difficult to measure. It is the cumulative effect of the improved service across the enterprise that resulted from the continuous coaching made possible by the direct and frequent feedback of the mystery shoppers on a daily basis. With a return on investment of nearly 100 percent, senior management concluded that this program was indeed profitable even before calculating the benefits in head count gained from the improved server performance.

Discussion

Using mystery shopping feedback for coaching employees incorporates well-researched psychological principles in that:

1. It is based on a job analysis of behaviors critical for effective server–customer performance. Hence, the relevance of feedback based on job analysis may decrease the likelihood that the relevance of the assessed behaviors will be questioned by employees, and increase the likelihood that “bottom-line” measures are affected favorably.
2. The feedback given to employees comes from a relatively neutral third party, a mystery shopper. This enables a manager to fulfill the role of coach with minimum concern that an employee will view the assessment of performance as a personal attack. A mystery shopper is a relatively “dispassionate observer.” Nothing is known about the employee by a shopper other than the service (performance) that is provided.
3. The feedback from a mystery shopper is provided shortly after observing an employee on the job. In the present study, it was 18 hours or less. Hence, the feedback is relatively timely.
4. The mystery shopper appears on a variable interval schedule. This enables coaching to occur on a systematic rather than a haphazard basis.

Theoretical Significance

The theoretical significance of the present investigation is at least threefold. First, consistent with reinforcement theory, feedback provided on a VI schedule increased and maintained high performance for both employees and organizations relative to a baseline condition. This is the first such finding regarding the effectiveness of providing feedback on a VI schedule in the workplace. Second, relevant to both goal setting and social cognitive theories, this study revealed that feedback does not need to be provided on a continuous schedule in order to maintain goal pursuit. Third, from the standpoint of theories on the criterion-space or workplace/job performance, this study suggests that a key driver of an organization's effectiveness in the service sector is the performance effectiveness of its frontline staff.

Practical Significance

The practical significance of the present findings for human resource management is at least fivefold. First, this study adds to the paucity of empirical research on the effect of coaching in an organizational setting. It is among the first to show that coaching can affect outcomes at both the individual and the organizational level of analysis. Whereas most coaching involves changing the behavior of upper-level managers, this study showed the benefits of a manager coaching lower-level employees who interact directly with the organization's customers.

Second, this is the first study to show the benefit of a manager coaching an employee on the basis of feedback from a relatively neutral third party, a mystery shopper, who knew nothing about an employee other than the service (performance) provided. Hence, ad hominem bias attributed by an employee to feedback from a supervisor was minimized.

Third, the benefits of using a mystery shopper intervention include the fact that it enables employee performance to be assessed on a VI schedule rather than continuously or at a fixed interval (e.g., annually, biannually, quarterly). Similar to students who are aware

that "pop quizzes" will be administered, employees do not know when an assessment of their performance will occur. Hence, a temporary spike in behavior that is typically observed on a fixed interval (i.e., studying only one or two nights before an exam; excelling on the job just prior to an appraisal) followed by a performance decrement subsequent to an exam/appraisal is likely avoided. When observations of one's performance are scheduled on a variable-interval rather than on a fixed-interval basis, ongoing excellence is required of employees in order for them to be assured that they will be observed performing effectively.

A fourth benefit of this intervention is that only one server was assessed each day. Consequently, a manager only had to spend a maximum of 30 minutes daily coaching an employee. This time requirement of a manager is far less than that which is required to coach all employees continuously.

Fifth, the results obtained in this study are likely generalizable to the service industry in general, and the hospitality industry in particular with regard to transfer of training to the job. These results suggest the benefit of coaching employees on a systematic variable-interval basis for ensuring that what is learned in a training program will continue to be applied on the job. Consistent with laboratory findings (Ferster & Skinner, 1957), performance in the present study increased when feedback was given on a variable interval schedule, and performance decreased when systematic coaching based on mystery shopping feedback subsequently ended. In the absence of this systematic coaching intervention, poor work habits were likely displayed to customers who had come to expect a high level of service. Consequently, customer count went down when employees were no longer being coached on the basis of feedback from a mystery shopper.

Whereas most coaching involves changing the behavior of upper-level managers, this study showed the benefits of a manager coaching lower-level employees who interact directly with the organization's customers.

Future research is now needed to see if procedures similar to that used in the present study can be adapted to other industrial sectors to facilitate transfer of training to the job.

Strengths, Weaknesses, and Subsequent Research

The primary limitation of this study is the use of a quasi-experimental design. It was not feasible to randomly assign servers to two or more conditions, nor was it possible to obtain a nonequivalent control group. Consequently, the data were analyzed using a cross-sectional time-series design and a GLS analysis. Specifically, observations were made prior to, during, and subsequent to the mystery shopping intervention in three restaurants. Few threats to internal validity appear to explain the close relationship between the introduction and removal of the mystery shopping intervention, and the parallel changes in a server's behavior and a restaurant's customer count (Campbell & Stanley, 1972). Cyclical maturation as a rival hypothesis for explaining the present findings was rejected because the mystery shoppers appeared at variable time periods of a day. Moreover, no obvious historical, extraneous events took place. For example, none of the restaurants initiated a new advertising campaign, changed managers, or instituted any "specials." Further, there were no internal management changes, nor were there any changes in the local economy in the time period in which this study took place. Employee tenure remained stable throughout this study. Finally, and arguably most importantly, the

results obtained in one restaurant were replicated in two additional restaurants.

Maturation as a rival hypothesis was also ruled out because we could assess the maturation trend prior to the introduction of the intervention. Instrumentation effects were also ruled out because there was no change in administration procedures regarding the ways company records were kept before versus after the introduction of the mystery shopper intervention. There was no change in menu items. Moreover, no server knew when a mystery shopper was present. Hence, the following conclusion appears warranted: Feedback from mystery shoppers, given to an employee by the manager/coach on a variable-interval schedule, increased both employee and organizational performance.

Nevertheless, because these findings are based on quasi-experimental design, the most worrisome causal alternative explanation of these results is that competent, aperiodic coaching, unassisted by mystery shopping data, might be as effective as the intervention used in this study.

Another arguable limitation of this study is that coaching is an applied technique that consists of myriad variables. There is no way of knowing which variable is more effective than another in bringing about a change in behavior. However, each variable has been shown empirically, in voluminous studies, to increase performance (e.g., reinforcement, feedback, goal setting) and, as Azrin (1977) noted, a set of variables when considered together can often yield a more effective treatment in an applied setting than any one of the variables alone.

GARY P. LATHAM is the Secretary of State Professor of Organizational Effectiveness in the Rotman School of Management at the University of Toronto. He is the recipient of both the Michael Losey award from the Society for Human Resource Management and the Herbert Heneman award from the Academy of Management's Human Resource Management Division. He is a past president of the Society for Industrial and Organizational Psychology as well as the Canadian Psychological Association. He currently serves on the Society for Human Resources Management (SHRM) board of directors. Latham is a fellow of the American Psychological Association, the Association for Psychological

Science, the International Association for Applied Psychology, the Academy of Management, the National Association of Human Resource Management, and the Royal Society of Canada.

ROBERT C. FORD (PhD, Arizona State) is a professor of management in the College of Business Administration (COBA) at the University of Central Florida (UCF), where he teaches management of service organizations. Dr. Ford has authored or coauthored numerous publications in both top research and practitioner journals, written five books, and serves on several editorial boards. He has served the Academy of Management (AOM) as editor of *The Academy of Management Executive*, director of placement, and division chair of two divisions. He has served the Southern Management Association (SMA) in every elective office, including president. He received the Distinguished Service Award from SMA and is an SMA fellow.

DANIEL TZABBAR is an assistant professor of entrepreneurship and strategy in the LeBow College of Business at Drexel University. He received his PhD from the University of Toronto, Rotman School of Management. His research focuses on the strategic implications of accessing and managing internal and external knowledge flow, and the facilitation of learning and technological change. Professor Tzabbar won the best dissertation award from the Technology and Innovation Management Division, Academy of Management Meeting, 2006. His research has appeared in *Academy of Management Journal*, *Strategic Organization*, *Advances in Strategic Management*, *Career Development International*, and *Industrial and Corporate Change*. He serves on the editorial board of *Group and Organization Management*.

References

- Ammons, R. B. (1954). Knowledge of performance, survey of literature, some possible applications and suggested experimentation, WASC Technical Report. 54-14, Wright Air Development Center.
- Andaleeb, S. S., & Conway, C. (2006). Customer satisfaction in the restaurant industry: An examination of the transaction-specific model. *Journal of Services Marketing*, 20, 3–12.
- Annett, J. (1961). The role of knowledge of results in learning: A survey. U.S. Navtradedevcen Technical Document report. 342-343.
- Ashford, S. J., & Northcraft, G. (2003). Robbing Peter to pay Paul: Feedback environments and enacted priorities in response to competing task demands. *Human Resource Management Review*, 13, 537–559.
- Azrin, H. H. (1977). A strategy for applied research: Learning based but outcome oriented. *American Psychologist*, 33, 140–149.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52, 1–26.
- Beck, J., & Miao, L. (2003). Mystery shopping in lodging properties as a measurement of service quality. *Journal of Quality Assurance in Hospitality & Tourism*, 4, 1–21.
- Bergh, D. D. (1993). Don't "waste" your time! The effects of time series errors in management research: The case of ownership concentration and research development spending. *Journal of Management*, 19, 897–914.
- Bowen, J., & Ford, R. C. (2002). Managing service organizations: Does having a 'thing' make a difference? *Journal of Management*, 28, 447–469.
- Campbell, D. T., & Stanley, J. C. (1972). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.
- Carrillat, F. A., Jaramillo, F., & Mulki, J. P. (2009). Examining the impact of service quality: A meta-analysis of empirical evidence. *Journal of Marketing Theory and Practice*, 17, 95–110.
- Chesbrough, H., & Spohrer, J. (2006). A research manifesto for services science. *Communications of the ACM*, 49, 35–40.
- DeStobbeleir, K., & Ashford, S. J. (in press). Feedback seeking behavior in organizations: Research, theory and implications. In R. Sutton (Ed.), *Handbook of criticism, praise, and advice*. New York, NY: Peter Lang.

- Dunnette, M. D. (1976). Mish-mash, mush and milestones in organizational psychology. In H. Meltzer & F. R. Wickert (Eds.), *Humanizing organizational behavior* (pp. 86–102). Springfield, IL: Charles C. Thomas.
- Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. East Norwalk, CT: Appleton Century Crofts.
- Finn, A. (2001). Mystery shopper benchmarking of durable-goods chains and stores. *Journal of Service Research*, 3, 310–320.
- Ford, R.C., & Bowen, D. (2008). A service dominant logic for management education: It's time. *Academy of Management Journal of Learning & Education*, 7, 224–243.
- Ford, R. C., Wilderom, C. P. M., & Caparella, J. (2008). Strategically crafting a customer-focused culture: An inductive case study. *Journal of Strategy and Management*, 1, 143–167.
- Gupta, S., McLaughlin, E., & Gomez, M. (2007). Guest satisfaction and restaurant performance. *Cornell Hotel & Restaurant Administration Quarterly*, 48, 284–298.
- Hall, D.T., Otazo, K. L., & Hollenbeck, G. P. (1999). Behind closed doors: What really happens in executive coaching. *Organizational Dynamics*, 27, 39–53.
- Heslin, P., Latham, G. P., & VandeWalle, D. (2005). The effect of implicit person theory on performance appraisals. *Journal of Applied Psychology*, 90, 842–856.
- Heslin, P. A., VandeWalle, D., & Latham, G. P. (2006). Keen to help? Managers' implicit person theories and their subsequent employee coaching. *Personnel Psychology*, 59, 871–902.
- Hsiao, C. (1986). *Analysis of panel data*. Cambridge, UK: Cambridge University Press.
- Hutcheson, P. G. (1996). Ten tips for coaches. *Training and Development*, 50, 15–16.
- Ilgen, D. R., Fisher, C. D., & Taylor, S. M. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, 64, 349–371.
- Komaki, J. L. (1994). Emergence of the operant model of effective supervision or how an operant conditioner got hooked on leadership. *Leadership and Organizational Development Journal*, 15, 27–32.
- Lance, C. E. (1994). Test of a structure of performance ratings derived from Wherry's (1952) theory of rating. *Journal of Management*, 20, 757–772.
- Latham, G. P. (2007). *Work motivation: History, theory, research and practice*. Thousand Oaks, CA: Sage.
- Latham, G. P., & Dossett, D. L. (1978). Designing incentive plans for unionized employees: A comparison of continuous and variable ratio reinforcement schedules. *Personnel Psychology*, 31, 47–61.
- Latham, G. P., & Locke, E. A. (2007). New developments in and directions for goal setting. *European Psychologist*, 12, 290–300.
- Latham, G. P., & Mann, S. (2006). Advances in the science of performance appraisal: Implications for practice. In G. P. Hodgkinson & J. K. Ford (Eds.), *International review of organizational and industrial psychology* (Vol. 21, pp. 295–337). Chichester, UK: Wiley.
- Lefkowitz, J. (2000). The role of interpersonal affective regard in supervisory performance: Literature review and proposed causal model. *Journal of Occupational and Organizational Psychology*, 73, 67–85.
- Lewis, R. C., & Chambers, R. E. (2000). *Marketing leadership in hospitality—Foundations and practices* (3rd ed.). New York, NY: Wiley.
- Liao, H. (2007). Do it right this time: The role of employee service recovery performance in customer-perceived justice and customer loyalty after service failures. *Journal of Applied Psychology*, 92, 474–489.
- Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35 year odyssey. *American Psychologist*, 57, 705–717.
- Luthans, F., & Peterson, S. J. (2003). 360 degree feedback with systematic coaching: Empirical analysis suggests a win-win combination. *Human Resources Management*, 42, 243–246.
- McDowall, D., McCleary, R., Meidinger, E. E., & Hay, R. A., Jr. (1980). *Interrupted time series analysis*. Thousand Oaks, CA: Sage.
- Paul, M., Hennig-Thurau, T., Gremler, D. D., Gwinner, K. P., & Wiertz, C. (2009). Toward a theory of repeat purchase drivers for consumer services. *Journal of the Academy of Marketing Science*, 37, 215–237.
- Pugh, S. D. (2001). Service with a smile: Emotional contagion in the service encounter. *Academy of Management Journal*, 44, 1018–1027.
- Sawyer, J. E., Latham, G. P., Pritchard, R. D., & Bennett, W. R., Jr. (1999). Analysis of work group productivity in an applied setting: Application of a time series panel design. *Personnel Psychology*, 52, 927–967.

- Schneider, B., Ehrhart, M. G., Mayer, D. M., & Saltz, J. L. (2005). Understanding organizational-customer linkages in service settings. *Academy of Management Journal*, 48, 1017–1032.
- Skinner, B. F. (1974). *About behaviorism*. Oxford, UK: Knopf.
- Smither, J. W., London, M., Flautt, R., Vargas, Y., & Kucine, I. (2003). Can working with an executive coach improve multisource feedback ratings over time? A quasi-experimental field study. *Personnel Psychology*, 56(1), 23–44.
- Strauss, J. P., Barrick, M. R., & Connerley, M. L. (2001). An investigation of personality similarity effects (relational and perceived) on peer and supervisor ratings and the role of familiarity and liking. *Journal of Occupational and Organizational Psychology*, 74, 637–657.
- Sue-Chan, C., & Latham, G. P. (2004). The situational interview as a predictor of academic and team performance: A study of the mediating effects of cognitive ability and emotional intelligence. *International Journal of Selection and Assessment*, 12, 312–320.
- Vander Wiele, T., Hesselink, M., & Van Iwaarden, J. (2005). Mystery shopping: A tool to develop insight into customer service provision. *Total Quality Management & Business Excellence*, 16, 529–541.
- Waldroop, J., & Butler, T. (1996). The executive as coach. *Harvard Business Review*, 74(6), 111–117.
- Watson, J. B. (1924). *Behaviorism*. New York, NY: Norton.
- Welch, J. (2001). *What I've learned from leading a great company and great people*. New York, NY: Headline.
- Wilson, A. M. (1998). The use of mystery shopping in the measurement of service delivery. *Service Industries Marketing*, 18(3), 148–163.