



Using Goals, Feedback, Reinforcement, and a Performance Matrix to Improve Customer Service in a Large Department Store

Nelson Eikenhout & JohnAustinPhD

To cite this article: Nelson Eikenhout & JohnAustinPhD (2005) Using Goals, Feedback, Reinforcement, and a Performance Matrix to Improve Customer Service in a Large Department Store, Journal of Organizational Behavior Management, 24:3, 27-62, DOI: [10.1300/J075v24n03_02](https://doi.org/10.1300/J075v24n03_02)

To link to this article: https://doi.org/10.1300/J075v24n03_02



Published online: 08 Sep 2008.



Submit your article to this journal [↗](#)



Article views: 1586



View related articles [↗](#)



Citing articles: 5 View citing articles [↗](#)

Using Goals, Feedback, Reinforcement, and a Performance Matrix to Improve Customer Service in a Large Department Store

Nelson Eikenhout
John Austin

ABSTRACT. This study employed an ABAC and multiple baseline design to evaluate the effects of (B) feedback and (C) a package of feedback, goalsetting, and reinforcement (supervisor praise and an area-wide celebration as managed through a performance matrix, on a total of 14 various customer service behaviors for a total of 115 employees at a large department store. In order to develop the intervention components in a manner that linked to organizational needs, an organizational functional assessment was first conducted using the Performance Diagnostic Checklist. The study occurred over a period of approximately 15 weeks, and the interventions were applied and evaluated across three departments, including cashiers, “Hardlines” sales, and “Softlines” sales. In all cases the behaviors occurred more often during the intervention phases than during either of the two baseline phases. The percent occurrence for all behaviors was also higher during the second baseline phase than during the first baseline. However, the percentages dropped noticeably for each measure during the second baseline after withdrawal of the feedback. The percentage point increase on the dependent measure scores from baseline 1 to the feedback (B) phase ranged from 30.7% for smiling in *Hardlines* to 64.9% for eye contact in *Softlines*. During the package in-

Nelson Eikenhout and John Austin are affiliated with Western Michigan University.
Address correspondence to: John Austin, PhD, Department of Psychology, Western Michigan University, Kalamazoo, MI 49008 (E-mail: john.austin@wmich.edu).

Journal of Organizational Behavior Management, Vol. 24(3) 2005

Available online at <http://www.haworthpress.com/web/JOBM>

© 2005 by The Haworth Press, Inc. All rights reserved.

Digital Object Identifier: 10.1300/J075v24n03_02

tervention phase (C), increases in the dependent measure scores ranged from 29.3% for small talk in *Softlines* to 49.9% for eye contact in *Hardlines*. Effect sizes were large in all cases, ranging from 1.58 to 2.42 for individual behaviors across conditions. Results are discussed in terms of the scale of the application, the effects of feedback and reinforcement in this and similar settings, and the use of a performance matrix as a management tool. [Article copies available for a fee from The Haworth Document Delivery Service: 1-800-HAWORTH. E-mail address: <docdelivery@haworthpress.com> Website: <<http://www.HaworthPress.com>> © 2005 by The Haworth Press, Inc. All rights reserved.]

KEYWORDS. Feedback, performance diagnostic checklist, praise, performance matrix, customer service, courtesy, department store

Today's economy is becoming increasingly service oriented. Businesses are finding it necessary to improve their level of customer service to survive in an increasingly competitive global market. The bestseller *Service America!* states:

The times have changed and we no longer live in a manufacturing economy. We now live in a very new economy, a service economy, where relationships are becoming more important than physical products. Just as America experienced an industrial revolution around the turn of the century, so we are now experiencing a service revolution. (Albrecht & Zemke, 1985, p. v)

Unfortunately, the amount of customer dissatisfaction with service has apparently also risen. "People are becoming much more conscious of the mediocre levels of service they receive in many aspects of their lives, and they are becoming mad about it" (Albrecht & Zemke, 1985, p. v). Whether performance levels among sales and service associates' have improved since that time or not, the ferocity and importance of service-based competition has, if anything, grown since the 1980's when *Service America!* was written. For example, more recently, J.D. Power and Associates released an industry study reporting that retail organizations have a competitive advantage if they can offer good value in products as well as an exceptional shopping experience for customers. Through a telephone survey, the study found that the most important factors for customers shopping at a value retailer were (in order of im-

portance): Value, sales and service associates, services, store environment, merchandise, reputation, sales and promotion, and store location (Anonymous, 2002). Further, during a recent documentary produced by *CNBC*, Wal-Mart executives reported estimating that losing a customer for life represents a loss of more than \$200,000 in revenue to their stores (*The Age of Wal-Mart . . .*, 2004). This estimate may be based on other reports in the area, such as the claim that it costs five times less to keep existing customers than it does to recruit new ones and that a 5% increase in customer retention is sometimes associated with a 25 to 85% increase in profits (*How to Measure . . .*, 2002). It is clear that service remains an important issue among U.S. business, and this is perhaps especially true in the retail industry.

There are many examples of successful applications of organizational behavior management (OBM) for manufacturing tasks (Stoneman & Dickinson, 1989), and other tasks in which a tangible, permanent output is produced such as applications processed (Wilk & Redmon, 1990). Studies dealing with behaviors that are more difficult to measure such as customer service, on the other hand, remain few in number compared to studies that deal with more tangible results. The chances that a manager might provide some feedback or reinforcement for a task in which an easily visible and lasting result is produced, such as stocking shelves, is higher than for customer service behaviors which are fleeting, difficult to measure, and the effects of which on customers may never be known. This is undoubtedly due to an absence of contingencies linking customer responses to quality of service and important consequences for managers on the sales floor. The traditional method for assessing customer service has been through customer satisfaction surveys. Unfortunately, these surveys often provide feedback that occurs long after the behavior responsible for the survey responses. In principle at least, the reinforcement value of delayed consequences may render them ineffective (Baum, 1994; Malott, 1992; Rachlin, 1989). In addition, getting customers to fill out survey forms can be difficult. Brown, Malott, Dillon, and Keeps (1980) handed out over 500 evaluation forms to customers in a department store, and even though there was a chance to win a \$100 gift certificate for returning a form, not a single evaluation form was returned. Sales figures may also be used as an indicator of overall customer service. However, since many other factors such as advertising, the season of the year, and economic conditions also affect and are, therefore, confounded with sales, sales cannot provide a direct measure of organizational consequences attributable to customer service. An al-

ternative to measuring customer service through surveys or sales is to measure through direct observation.

A difficulty with directly observing customer service is defining the exact appearance (i.e., topography) of outstanding customer service behavior and making a person available to measure it when it occurs. Brown et al. (1980) operationalized customer service by isolating four classes of operant behavior: approach, greeting, and courtesy to each customer and closing customer sales. Each component was described in a specific operational definition that was then used to collect data concerning the frequency of each behavior via direct observation. An intervention, combining training and feedback, improved customer service behavior among three employees. In another example, Luthans, Paul, and Taylor (1986) developed a list of 14 functional and three dysfunctional behaviors for salespersons in a department store setting. These behaviors included both customer service and non-service behaviors. Participants who met or exceeded established performance standards received some time off from work and the chance to win vacation plane tickets. As a result, functional behavior increased and dysfunctional behavior decreased. Changes in frequencies of specific functional behaviors, such as those related to customer service, were not reported.

Performance feedback and reinforcement have been the most commonly used OBM interventions to improve customer service. Brown and Sulzer-Azaroff (1994) used feedback to improve several customer service behaviors (smiling, greeting, and looking at the customer) among three bank tellers. Welsh, Bernstein, and Luthans (1992) used an intervention based on the Premack principle to improve a variety of behaviors related to customer service in a restaurant setting among eight employees. Komaki, Blood, and Holder (1980) used an intervention combining behavioral cues and social reinforcement to successfully improve the frequency of smiling at customers in a fast food restaurant among 11 employees.

Most of the studies attempting to improve customer service have used individual level interventions quite successfully. However, given the difficulty and time required to acquire an adequate sample of each employee's behavior as the basis for delivering feedback and reinforcement, most studies have restricted sample sizes to fewer than 15 employees. Although even small scale demonstrations are important, if the cost is too high to use the intervention among larger numbers of employees, the logical consequence is that the overall impact may be less than desirable. As the number of employees in an organization in-

creases, so does the practical need for some type of strategy for customer service improvement at the group level of implementation and analysis. This study used a group-level strategy to effect improved customer service behavior among more than 110 employees in a single retail store.

Since customer service consists of at least several behaviors such as greeting, smiling, and eye contact, feedback and reinforcement must be delivered for each behavior individually or for some outcome measured by combining observations of all service-related behaviors targeted for improvement. Just as the need for group-level interventions is positively correlated with workforce size, so too is the need for some global, or summary, measure of customer service related behaviors. A few studies have provided consequences based on a global customer service score. Welsh et al. (1992) used a weighted checklist as a tool to provide reinforcement for a variety of customer service behaviors to employees in a fast food restaurant. Crowell et al. (1988) measured customer service for bank tellers using 11 verbal behavioral targets. Quality point values were assigned to the 11 behaviors targeted for improvement so that the total possible points that could be earned in any transaction was 100.

Although its efficacy has not been documented in the research literature, another similar matrix method used to provide feedback and reinforcement for measures on a number of behaviors, results, and/or ratings was described by Daniels (1989). This version of the performance matrix, described in Daniels (1989), evolved from the objectives matrix (OMAX) developed by Riggs (1986) in Rigg's attempt to measure "tender loving care" in a hospital setting. It should be noted, however, that there are many potential incarnations of a performance matrix that can be used to improve management of organizational performance including those described by Gilbert (1978), Abernathy (2001), and Kaplan and Norton (1992). The current study made use of a performance matrix most similar to the one described by Daniels but not unlike those used in the Welsh et al. (1992) and Crowell et al. (1988) studies.

The goal of the current study was to improve a variety of customer service behaviors storewide in a large department store using direct observation and an intervention package developed from a functional assessment procedure. The intervention package consisted of goal setting, group performance feedback, immediate social reinforcement, and tangible reinforcement. The tangible reinforcement delivery was based on points earned in accordance with rules specified by and reported in a

performance matrix during each of several observation periods for three store areas. Also, this study constituted an attempt to extend the research on customer service by using a group-level strategy covering a larger number of employees than had been covered in previous studies.

METHOD

Setting and Participants

The setting for the study was a large department store located in the Midwest United States. The store was divided into several main areas: the checkout lanes, the restaurant, the service counter, and the general sales floor which was subdivided into two main areas, the *Hardlines* and the *Softlines*. The *Hardlines* area included electronics, pet supplies, health and beauty, candy and snacks, cards and party items, home office, home floral, books, music and movies, toys and games, seasonal items, luggage, sporting goods, home storage, cleaning supplies, hardware, and automotive. The *Softlines* area included jewelry, cosmetics and accessories, men's wear, women's wear, sportswear, shoes, hosiery, intimate apparel, infant and children's clothing, domestic items, housewares, gifts, lighting, furniture, and small electronics. Customer service behavior was measured for all employees on the general sales force and at checkout lanes, but not at the service counter or restaurant.

One hundred and fifteen employees served as participants in this study. This group consisted of all the employees working as cashiers (45) and all the salespersons on the general sales floor (70). Half of the salespersons (35) worked in the *Hardlines* area and the other half (35) worked in the *Softlines* area. Employees were assigned to one of the three areas when first hired and were not rotated from one store area to another. The only exception was that about half of the employees working on the general sales floor were also trained as backup cashiers, and occasionally helped assist customers in the checkout lanes when the store was busy. Ten to fifteen employees were assigned to work in each area at a given time during store hours. Approximately 35 of the employees worked full-time. The remaining employees worked, on average, from 12 to 30 hours per week. All employees earned an hourly wage and punched in and out on a time clock each day they worked.

DEPENDENT VARIABLES

Poor customer service was described as a source of concern during an early meeting with the store manager. The chain of department stores had developed a set of criteria to help identify good customer service and called it the “B.G.O.R.” model, with “B” standing for “be friendly,” “G” for “greet the customer,” “O” for “offer assistance,” and “R” for “respond quickly.” The behaviors pinpointed as constituting good customer service were based on this model, and were an attempt to link the intervention to an overall organizational goal. The store manager indicated that he was interested in having data that demonstrated an improvement in customer service related to the “B.G.O.R.” model because it would help the corporate managers view him and the entire store more positively. He also indicated that he was motivated to participate in the study because if the customer service behaviors improved as a result of the intervention, it would help him distinguish himself and the store from other stores in their chain. Therefore, measures of the behaviors that served as dependent variables for this study, described below, were related to the store manager’s operational and career goals and objectives.

Customer Greeting

Customer greeting was defined as saying such things as “hello,” “good morning,” or “how are you doing today?” to any customer that passed by (within two meters) or was shopping in any area in which an employee was working. The measure was expressed as a percentage, e.g., the number of times employees greeted customers out of all customer greeting opportunities observed during an observation session. A customer greeting opportunity was defined as an instance when a customer passed by an employee and that customer was not talking to another person or facing away from the employee, or an instance when a customer was shopping (i.e., inspecting products) in the area an employee was working.

Assistance Offered

Offering assistance was defined as saying such things as “are you finding everything you need O.K.?” or “may I help you with something?” to any customer who was shopping in the area where the employee was working, assuming the customer had not been assisted by another employee. Assistance offered was also expressed as a percent-

age, or number of times the behavior occurred divided by the number of opportunities for its occurrence.

Smiling

This was defined as a percentage, or number of observed customer-employee interactions during which the employee was observed smiling. Any employee smile directed at a customer during any part of their interaction, for any duration, was scored as an occurrence for purposes of computing the raw score of smiling.

Eye Contact

Eye contact was defined as a percentage, or number of times employees were observed looking at and orienting toward a customer who passed by or who was shopping in the area in which an employee was working divided by the total number of opportunities for making eye contact. In order to obtain reliable measures of eye contact, only employees who were observed to be facing or turning toward the customer and looking in the direction of the customer's face were counted as having made eye contact. Measuring an occurrence of eye contact did not require that the customer also look at the employee.

Small Talk

Small talk was measured as a percentage, or number of times employees were observed talking to customers about things other than just assistance with what they were looking for or for what customers were purchasing divided by total number of opportunities for such interactions. Small talk could occur at any time employees had the opportunity to greet or assist customers and the customer did not walk away from the employee within five seconds after being in the employee's area. When the content of a conversation between employees could not be heard by the data collectors, no data regarding small talk were recorded for the interaction.

Several other behaviors were only measured during baseline. These were responding to customer requests, prompting for accessories, acknowledging waiting customers, time to respond verbally to a request made through a walkie-talkie or through the intercom, and the time to meet a customer after a remote request had been made. The data collected for these behaviors were considered too variable for further anal-

ysis. That is, they presented too few observation opportunities (in the cases of acknowledge waiting customers, prompting for accessories, and time to meet a customer from a remote request) or they were at too high a level during baseline (in the cases of responding to requests and time to respond verbally to a remote request) to warrant targeting them for change during the intervention. In addition, the store manager indicated a preference that the interventions focus on the five target behaviors described above. This preference appeared to be in response to some of the baseline data.

All five behavioral measures of customer service were collected concurrently and separately in the *Hardlines* and the *Softlines* areas, and all the measures except assistance offered were also collected concurrently and separately at the checkout lanes during the study.

OBSERVATION PROCEDURES AND INTEROBSERVER AGREEMENT

The first author collected the dependent measures of customer service three times a week at random during store hours. The observation days and times were purposefully varied each week to obtain a fair sampling of behavior for as many employees as possible. Employees were not informed that their behavior was being observed during baseline. Nevertheless, a few employees remarked that they had noticed the primary observer in the store, but that they thought he was working to control shoplifting. A minimum of 20 customer interaction opportunities was recorded for each dependent variable in each area of the store (*Hardlines*, *Softlines*, and *Checkout*) during each session of data collection. This requirement for a minimum number of observations was set to ensure a fair sampling of behavior in each store area and to control for unnecessary between-session variability due to behavior samples of different sizes. Also, observations were rotated from *Hardlines* to *Softlines* after each customer interaction opportunity so that the observers would not seem to be lingering in one store area or focusing on a particular employee. Observations for *Checkout* were collected in at least two blocks per session. Again this was intended to limit the possibility that employees would conclude that the observers were watching them. Observation sessions ranged from two to three hours in length with a mean session length of two hours, ten minutes.

A second observer collected data separately, along with the primary observer, on one full session randomly determined out of the three ob-

observation sessions per week. Therefore, there was an interobserver reliability check for approximately 33% of the data collected. The second observer was trained by the first observer prior to the first interobserver reliability check session. During training, the first observer reviewed the definitions of the behaviors with the second observer. Next, the first observer had the second observer accompany the first observer during an actual session. The training session consisted of both observers performing at least 20 observations together per store area. After each observation the first observer reviewed and discussed the data sheet with the second observer. The data from the second observer during the training session are not included in the results for interobserver agreement. Observations by the second observer were conducted simultaneously, but independently from the primary observer. However, whenever a customer service opportunity was noticed by either observer, that observer would signal the start of an independent observation. Both observers then recorded the time of the observation on the data sheet. The second observer was also naïve to the purpose of the study and to phase changes.

An interobserver reliability score was computed by adding up the total number of agreements for the behaviors observed, dividing by the total number of agreements and disagreements during the session, and multiplying by 100.

FUNCTIONAL ASSESSMENT

An organizational functional assessment was performed during the first baseline phase to determine aspects of the environment that may have been supporting or hindering employees from engaging in customer service behaviors (Austin, Carr, & Agnew, 1999; Austin, Weatherly, & Gravina, in press). Interviews held during weekly meetings with the store manager and occasional meetings with departmental managers along with notes taken during direct observation of the antecedents and consequences recorded during customer interactions were used as the basis for the functional assessment. The performance diagnostic checklist (PDC) (Austin, 2000) that included questions about the antecedents and information, equipment and processes, knowledge and skills, and consequences was completed during interviews with the store and departmental managers, and when appropriate, during direct observations.

All employees had been introduced to the organization's mission, job mission, and objectives and during a training program that included CD-

based and peer-based training shortly after being hired. The CD-based training included video demonstrations of examples of good and poor customer service that were based on the store's "B.G.O.R." model described earlier. No instances of demonstrating or discussing good customer service were observed when the first author directly observed peer training sessions. No goals related to customer service had been set by the organization, although goals were used for behaviors that produced more tangible results such as stocking shelves and cleaning areas of the store. Several competing behaviors were identified during the assessment. First, some employees spent much of their time stocking shelves which, in some instances, might have been an attempt to avoid dealing with customers since an employee stocking shelves was rarely observed facing or looking at customers in the area. Employees were, however, not provided with an incentive for stocking shelves or for sales. When asked what the consequences were for stocking shelves, the store manager said that when an employee failed to stock shelves correctly it was easy for a manager or supervisor to find out and talk to the employee. Second, employees were commonly observed talking to coworkers and many employees would walk together through the store in pairs or in small groups. Also, employees were often observed turning and walking up an adjacent aisle as customers approached. This may have indicated some response effort associated with interacting with customers for some employees. There were no observed obstacles in the work environment or problems with equipment that could have prevented employees from emitting the customer service behaviors measured.

The lack of behavioral consequences was found to be the area of greatest deficiency during the functional assessment. No performance feedback was provided to employees during the baseline conditions for any of the pinpointed customer service behaviors. In addition, although supervisors spent at least two to three hours per shift out on the sales floor with employees, not a single observed instance of feedback or social reinforcement delivered to employees by supervisors was observed during the first baseline period of the study. The store manager indicated that, in the past, when a customer had praised the service provided by a particular employee, the employee was praised in turn by the manager. The store manager indicated that this type of reinforcement for good customer service was very rare and had occurred at most about two to three times a month for an exemplary employee. The store manager also reviewed customer satisfaction survey results with employees and used the results to praise employees. There appeared to be no other types of reinforcement available to employees who engaged in good customer service behavior, other than those provided by

customers directly, and this included no tangible reinforcement. The store manager also indicated that when customer complaints occurred, either he or the department managers would speak directly with the employee involved in the complaint. According to the store manager, customer complaints were uncommon. He indicated that most employees responded to customer requests effectively. His concern was that employees were not initiating contact with customers frequently enough.

PROCEDURES AND EXPERIMENTAL DESIGN

A within-groups ABAC research design was used to evaluate the effects of two independent variables on five dependent customer service related employee behaviors across $N = 3$ employee groups in three store areas. Intervention I (Phase B) included group-by-group posted graphic raw scores (three times per week) and was followed by a reversal to baseline phase (A) in which all graphs were removed and employees were informed by their supervisors that the graphs would no longer be posted. The Intervention Package II (Phase C) replicated the first intervention, graphic feedback, plus introduction of the performance matrix that presented initial raw scores, performance goals, weights, and, on a weekly basis, points earned per week by each store area group. In addition, this package included immediate positive social remarks by the store manager and area managers whenever they observed occurrences of target behaviors among members of the work force. Finally, the Intervention Package II (Phase C) included a contingency between raw scores earned per store area and access to a single store area celebration party. That is, only employees from the store areas that scored at least 1000 points on the performance matrix during the final two weeks of the phase were permitted to attend an area store party.

Prior to collecting data, the experimental design and procedures were described to the store manager. He accepted having the graphs removed, as long as the reversal to baseline was no longer than two to three weeks. Also, the study was conducted at no cost to the store, which no doubt contributed to the store manager's tolerance of the reversal.

Baseline (Phase A)

The first baseline phase lasted approximately six weeks. Baseline data were collected through direct observations of the dependent variables. Although employees may have seen the experimental observers frequenting the store floors, employees were never informed (during the entire study) of the variables being measured by observers.

Intervention I: Graphic Feedback (Phase B)

Next, graphic feedback for customer greeting only was implemented across the three store areas (i.e., *Hardlines*, *Softlines*, and *Checkout*). Some spread of the effects was observed, especially for eye contact and smiling, and therefore, on the third observation session after beginning the graphic feedback for customer service, graphic feedback was also provided for the remaining dependent measures.

Three times each week during this intervention period, graphs of performance for each targeted behavior were posted on a wall near the time clock where employees punched in and out. Therefore, all employees could easily see the graphs each day they worked. The graphs depicted each group's raw scores for each target behavior across each observation session which included all prior sessions from the Baseline (phase A) up to the most recent observation session. In addition, a dotted line separated the baseline data from data collected when the graphic feedback was posted. No mean lines per week or by phase were ever posted on the graphs. The graphs posted appeared as they are depicted in the results section of this article except that dates for which each observation was conducted appeared where the session numbers are located on the charts following (Figures 1, 2, and 3). The graphs were posted separately for *Hardlines*, *Softlines*, and *Checkout*, but all were posted in the same area next to the time clock. Each graph was clearly labeled with the behavior being measured, the dates the measures were collected, the date the graph was posted, and the area of the store (i.e., *Hardlines*, *Softlines*, or *Checkout*) where data for the measures had been collected. In addition, each graph contained a verbal description of what was being measured and an operational definition of what each behavior "looked like" when it occurred and was observed. These graphs of performance were also posted three times a week during Intervention Package II (Phase C) in the same manner as in Intervention I (Phase B). The only difference with the graphs posted during Intervention Package II (Phase C) was that the upcoming week's goals were also depicted on the graphs as a straight line drawn across the area where the week's data points would be placed. Thus, the Intervention Package II (Phase C) consisted of the first Intervention Package (Phase B) plus goal setting, additional feedback as points earned via the performance matrix, social recognition for targeted behavior provided by the store manager and area manager, and the celebration party contingency.

FIGURE 1. Percentage of Observations in Which Target Behaviors Occurred for *Hardlines*

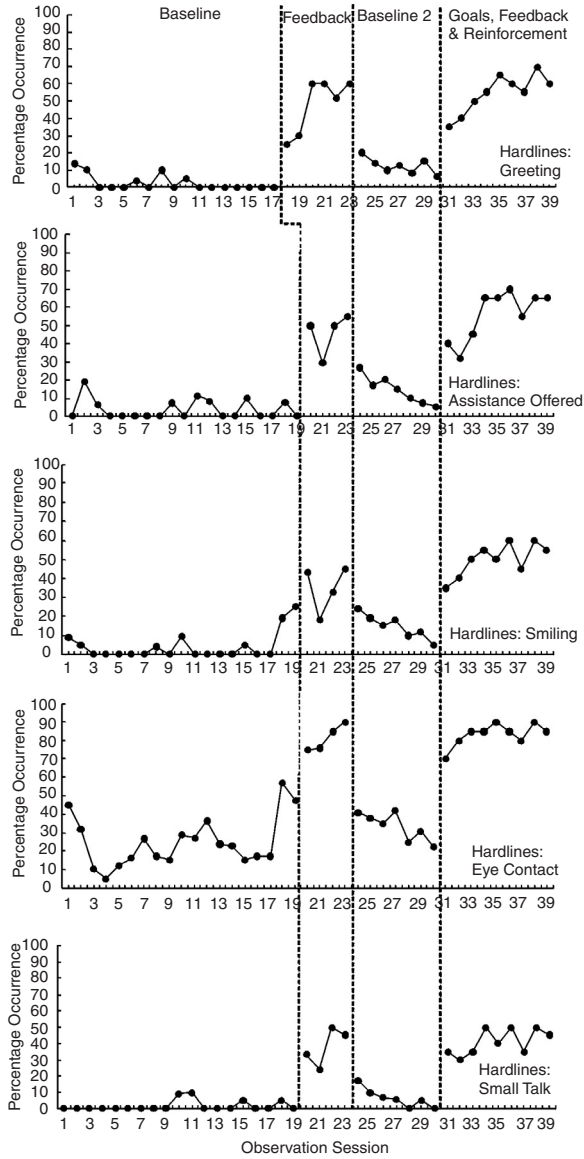
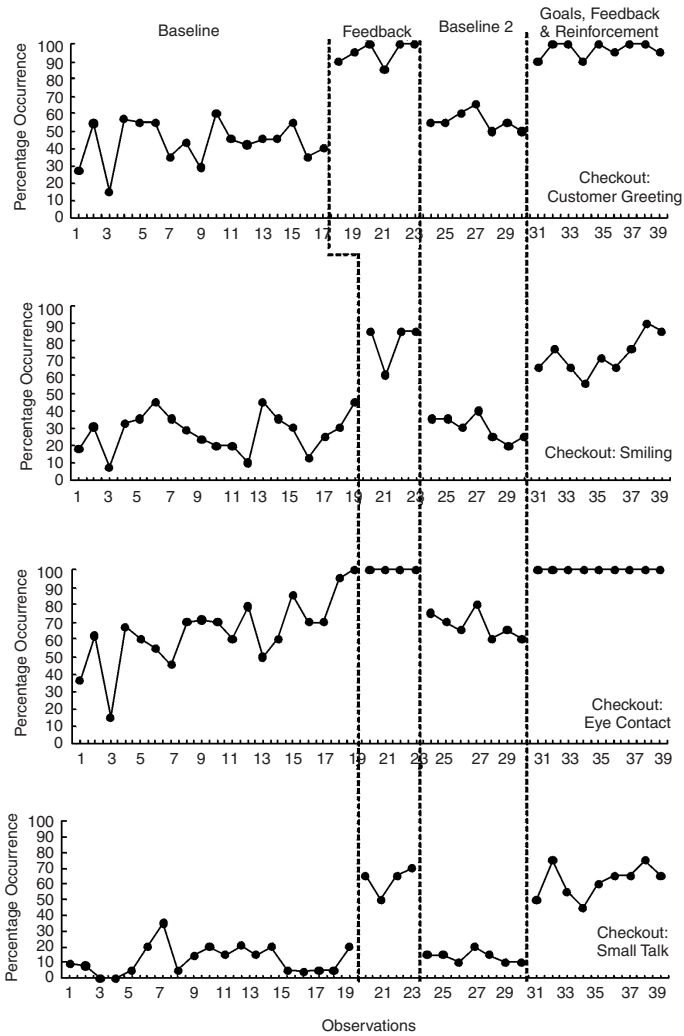


FIGURE 3. Percentage of Observations in Which Target Behaviors Occurred for Checkout



Reversal to Baseline (Phase A)

During this phase graphs of performance were not posted for *Hardlines*, *Softlines*, and *Checkout*. Data were collected through direct observation of the dependent variables. Employees were not told why the graphs were no longer being posted, and if employees questioned su-

pervisors regarding the graphs, supervisors were instructed to say that the store manager had decided to stop posting the graphs, but that they did not know why graphs were no longer being posted. This phase lasted for two weeks.

Intervention Package II (Phase C)

This phase began by implementing an intervention package of weekly goals, weekly feedback and social reinforcement for improvements in performance, immediate feedback and supervisor praise for employees “caught in the act” of good customer service behavior; and a celebration contingency for meeting an overall ultimate customer service goal. Each component of the intervention was communicated to employees verbally by the store and department managers formally during weekly team meetings, and also informally during supervisor interactions with subordinates. Each of the intervention components is described in more detail below.

Weekly Goals. At the beginning of the Intervention Package II (Phase C), the first author met with the store manager and reviewed the average percentages for each behavior measured during the Reversal to Baseline phase. At this meeting performance goals for week one and the ultimate goals for weeks two and three of the Intervention Package II (Phase C) were assigned by the store manager with input from department managers for each of the targeted behaviors for each store area using the performance matrix. The store manager and department managers were also trained on how to use the performance matrix during this meeting. The purpose for using the performance matrix was to communicate to employees the weekly goals and performance feedback in the form of an overall point score that summarized performance on all measured customer service targets. In other words, the matrix was used to inform employees how well each store area performed each week in relation to the performance goals set. The first week’s goals were set approximately halfway between the mean performance scores of the second baseline phase and the final, or ultimate goals, stated on the performance matrix. The first week’s goals appear in Figures 4, 5, and 6, and correspond to column seven on each store area’s performance matrix. The purpose for the goals during the first week was to provide a basis for performance feedback and reinforcement from managers that would hopefully increase the probability that the store areas would achieve the higher, ultimate goals for the next two weeks. The goals for the second week corresponded to column 10 and reflected ultimate goal attainment

FIGURE 4. Performance Matrix for Week 3 of Phase 4 for *Hardlines*

PERFORMANCE MATRIX

Store Area: Hardlines

Date:

<i>Behaviors (measured in %)</i>	4	5	6	7	8	9	10	11	12	Weight	Raw Score	Points
<i>Greeting Customers</i>		15		40		50	60	70	80	30	62	300
<i>Offering Assistance</i>		20		40		50	60	70	80	30	62	300
<i>Smiling</i>		15			30	40	50	55	65	15	55	165
<i>Eye Contact</i>		30		60		80	90	95	99	15	85	135
<i>Small Talk</i>		10			25	35	40	50	60	10	43	100
	Current					GOAL					Total Score: 1000	

Reinforcement Plan:

<i>POINTS</i>	<i>R+ Criteria</i>	<i>Comments</i>	<i>Plans</i>
1000	Meet for 2 consecutive weeks		Store Celebration

FIGURE 5. Performance Matrix for Week 3 of Phase 4 for *Softlines*

PERFORMANCE MATRIX

Store Area: Softlines

Date:

<i>Behaviors (measured in %)</i>	4	5	6	7	8	9	10	11	12	Weight	Raw Score	Points
<i>Greeting Customers</i>		15		40		50	60	70	80	30	60	300
<i>Offering Assistance</i>		20		40		50	60	70	80	30	63	300
<i>Smiling</i>		15			30	40	50	55	65	15	63	165
<i>Eye Contact</i>		30		60		80	90	95	99	15	90	150
<i>Small Talk</i>		10			25	35	40	50	60	10	45	100
	Current					GOAL					Total Score: 1015	

Reinforcement Plan:

<i>POINTS</i>	<i>R+ Criteria</i>	<i>Comments</i>	<i>Plans</i>
1000	Meet for 2 consecutive weeks		Store Celebration

FIGURE 6. Performance Matrix for Week 3 of Phase 4 for *Checkout*

PERFORMANCE MATRIX

Store Area: Checkout

Date:

<i>Behaviors (measured in %)</i>	4	5	6	7	8	9	10	11	12	<i>Weight</i>	<i>Raw Score</i>	<i>Points</i>
<i>Greeting Customers</i>		55		70	80	85	90	95	100	40	98	440
<i>Smiling</i>		30		50	65	80	90	70	90	20	83	180
<i>Eye Contact</i>		70		80	90	95	100			30	100	300
<i>Small Talk</i>		10		25	40	50	60	70	80	10	68	100

Current

GOAL

Total Score: 1020

Reinforcement Plan:

<i>POINTS</i>	<i>R+ Criteria</i>	<i>Comments</i>	<i>Plans</i>
1000	Meet for 2 consecutive weeks		Store Celebration

as indicated on the performance matrix. The third week’s goals were set at the same level as the second week’s goals. Goals were not raised during the third week because goal attainment for the last two weeks of the Intervention Package II (Phase C) represented an acceptable level of performance to management, and attendance at the celebration was designed to be contingent on employee performance over the course of two weeks. The goals were identified clearly by a line drawn on the graphs in the area where the three data points for the upcoming week would be plotted. In addition, the performance matrices for each store area that depicted the ultimate goals and weights for each behavior were placed each week next to the graphic feedback near the time clock. These ultimate goals and behavioral weights which were used to determine if one or more store areas would qualify for a celebration were unchanged on the matrices posted during the three weeks of Intervention Package II (Phase C). The data shown on the graphic feedback, however, remained dynamic and was posted before each observation session three times a week. The graphic feedback depicted all raw scores for each session since the beginning of the first baseline phase and in-

cluded lines drawn to separate phase changes. How the goals on the performance matrix were determined and how to read and interpret the matrix, as well as the consequences for achieving the goals were communicated to employees in *Hardlines*, *Softlines*, and the *Checkout* lanes by the department managers before Intervention Package II (Phase C) began. The upcoming week's performance goals were also communicated to employees during the weekly team meetings all employees were required to attend.

In the performance matrix (refer to Figures 4-6 for examples) used in this study behaviors targeted for improvement appeared as row headings in the first column and point values (from 4 to 12 points) appeared as column headings for the next nine columns. The remaining column headings were, respectively, *Weight*, *Raw Score*, and *Points* (for points earned per targeted behavior). The *Weight* values, one per target behavior, were required to sum to 100. *Raw Score* referred to the percentage of observations collected during an observation period for which each target behavior was observed to occur. Raw scores were calculated by computing the average percentage of observations for which each behavior occurred during each of the three observation sessions each week. Percentage, in this context, referred to the number of times a target behavior was observed to occur divided by the total number of observations of behavior or opportunities for the behavior that occurred during that particular observational period or session. Within the cell at the intersection of each row (each target behavior was in a different row) and point value column (columns 4 through 12) a performance goal level could be specified. However, for the matrices used in this study, no performance goals appeared in columns 4 or 6 for the three store areas. Column 5, on the other hand, included a performance which was the mean from the entire reversal to baseline phase for each of the target behaviors (i.e., greeting customers, offering assistance, smiling, eye contact, and small talk) for each of the three store areas. The ultimate goal appeared in column number 10 for each of the targeted behaviors for each store area. This goal varied across store areas, depending on baseline performance of the targets. Therefore, as described above and as shown in Figures 4-6, the points columns between 5 and 10 specified intermediate goal levels for which points could be earned. Columns 11 and 12 for all store areas included performance goals that were higher than the ultimate goals in each area's column 10 goal column.

The matrix integrated these various performance goals with point values to arrive at points earned per target behavior per store area. The

first step in operationalizing feedback using the matrix occurred when an author filled in newly collected data into the *Raw Score* column for each target behavior. When calculating the point value of a behavior in the matrix, an author located the columns depicting performance goals just above and just below the newly observed raw score. For example, if the raw score for greeting customers was observed to be 62 and columns 10 (ultimate goal) and 11 (overachievement of the ultimate goal) had fixed raw score goals of 60 and 70 assigned to them, then the points earned for that raw score corresponded to column 10 because the raw score exceeded the goal of 60, but did not exceed the goal of 70. After the column corresponding to the observed raw score was located, points for each behavior were calculated by multiplying that column heading number by the weight given to that behavior. Therefore, using the example above, if the weight assigned to greeting customers was 30, the points earned for that behavior would have been 300 (weight of 30 \times column 10). Finally, the scores for each behavior entered in the right most column of the matrix (i.e., the *Points* column) were summed to get a total performance score for the store area. The matrices used in this study were constructed so that ultimate goal attainment across all behaviors in the matrix for each store area resulted in a score of 1000 points for each store area. Once goals for the matrix were filled out by managers they remained constant during the three weeks of the Intervention Package II (Phase C). These goals were depicted in column 10 on the three performance matrices. Scores on the matrix were used as the basis for delivering a single area-wide consequence (i.e., banquet and celebration).

Weekly Feedback and Social Reinforcement for Improvements in Performance. At least once each week (i.e., during the team meeting) the supervisor for each store area gave feedback on customer service measures and congratulated employees for improvements in customer service performance, and especially for goal attainment. The store manager and the managers for the three store areas wrote positive comments on the posted graphs at least once a week contingent on improvements in performance, but often this occurred as frequently as three times a week. Also, as a check of independent variable integrity, managers marked a check on a data sheet at each team meeting in which reinforcement was delivered.

Immediate Feedback and Social Praise for Employees "Caught in the Act" of Good Customer Service Behavior. All store supervisors were encouraged to actively look for instances of good customer service behavior when they were on the store floor, and to praise employees for it while on the floor. The supervisors were trained by the first author on how to provide

the feedback. The feedback described the appropriate behavior that was observed followed by a statement of praise and encouragement. Supervisors were instructed to walk through their store area at least three times a day with the sole purpose of praising appropriate customer service behavior. In order to ensure consistency as well as immediacy of the reinforcement, supervisors were also instructed not to ignore any instances of good customer service observed by employees. Supervisors observed delivering reinforcement to employees on the store floor and at checkout lanes were also praised for it by the primary observer and store manager. Supervisors were required to keep a log of reinforcement they delivered on the store floor. This log was brought to a weekly meeting with the store manager and the first author. During these meetings, the store manager praised the supervisors for completing the reinforcement logs.

Celebration Contingency for Meeting an Overall Customer Service Goal. A contingency was described to employees wherein a celebration would be held based on the points calculated by the performance matrix if customer service in any of the three store areas reached a final goal level as determined by the store supervisor. In addition to setting weekly goals for each behavior, the store manager (with input from the department managers), set an overall customer service goal for each store area. The overall customer service goal for each store area was 1000 points as scored on the performance matrix during week two and week three of the Intervention Package II (Phase C). If customer service in a store area met or exceeded that goal for the final two weeks of the Intervention Package II (Phase C), then all employees working in the area or areas that met the goals were eligible to attend an area party celebrating exemplary customer service. The single celebration was to be free for all employees from the store areas *that qualified* and included food and beverages. That meant that if only one or two store areas qualified for the celebration, employees from the areas that did not qualify would not be able to attend. During all three weeks of the Intervention Package II (Phase C) the goals were communicated by department managers at the weekly meetings to employees and indicated on the posted graphs and on the performance matrices.

RESULTS

Table 1 shows the average, range and SD of dependent measures by phase for the three store areas observed. Effect size (d) was calculated for each treatment phase and the baseline phase that immediately pre-

TABLE 1. Mean, Standard Deviation, and Range for All Dependent Variables and All Phases

Store Area/Behavior	Baseline 1	Feedback	Baseline 2	Goals, Fdbk & R+
HARDLINES				
Customer greeted average	2.5	47.8	12.3	54.4
<i>Standard Deviation</i>	4.47	16.13	4.73	11.3
<i>Range</i>	13.3	35	14	35
Assistance offered average	3.7	46	14.3	55.8
<i>Standard Deviation</i>	5.47	11.58	7.65	13.57
<i>Range</i>	18.8	26	21.7	38
Smiling average	4	34.7	14.7	50
<i>Standard Deviation</i>	7.13	12.32	6.27	8.66
<i>Range</i>	25	27	18.8	25
Eye contact average	24.8	81.5	33.4	83.3
<i>Standard Deviation</i>	13.76	7.23	7.75	6.12
<i>Range</i>	52.1	15	20	20
Small talk average	1.5	38	6.4	41.1
<i>Standard Deviation</i>	3.16	11.75	5.99	7.82
<i>Range</i>	9.5	26	17.4	20
SOFTLINES				
Customer greeted average	2.2	57	12.1	55
<i>Standard Deviation</i>	4.74	14.35	5.86	10.61
<i>Range</i>	17	40	16.8	35
Assistance offered average	4	55.8	17.3	56.1
<i>Standard Deviation</i>	6.2	21.93	7.67	13.18
<i>Range</i>	18	50	19.4	40
Smiling average	4.9	54.7	13.7	58.3
<i>Standard Deviation</i>	6.35	20.19	4.83	10.61
<i>Range</i>	18.8	43	14.7	30
Eye contact average	23.4	88.3	39.2	84.4
<i>Standard Deviation</i>	12.17	8.06	8.23	9.5
<i>Range</i>	45	16	24.5	25
Small talk average	3.8	36.1	11.9	41.2
<i>Standard Deviation</i>	5.07	14.67	4.42	6.87
<i>Range</i>	15	32	13.2	20

TABLE 1 (continued)

Store Area/Behavior	Baseline 1	Feedback	Baseline 2	Goals, Fdbk & R+
CHECKOUT				
Customer greeted average	38.8	95	55.8	96.7
Standard Deviation	12.31	6.32	5.35	4.33
Range	45	15	15	10
Smiling average	28	78.8	30	71.7
Standard Deviation	11.16	12.5	7.07	10.9
Range	37	25	20	35
Eye contact average	64.2	100	67.9	100
Standard Deviation	19.71	0	7.56	0
Range	85	0	20	0
Small talk average	11.9	62.5	13.6	61.7
Standard Deviation	9.13	8.66	3.78	10.31
Range	35	20	10	30

ceded it, and these data are presented in Table 2. Effect size (d) was calculated by subtracting the treatment mean from the baseline mean, and dividing by the pooled standard deviation of the two phases. In all cases the behaviors occurred more often during the intervention phases than during either of the two baseline phases. The percentages were also higher during the second baseline phase than during the first baseline. However, the percentages dropped noticeably for each measure during the second baseline after withdrawal of the feedback. The percentage increase on the dependent measure score from baseline 1 to the feedback phase ranged from 30.7% for smiling in *Hardlines* to 64.9% for eye contact in *Softlines*. During the Intervention Package II (Phase C), increases in the dependent measure scores ranged from 29.3% for small talk in *Softlines* to 49.9% for eye contact in *Hardlines*.

Figure 1 depicts the percentage of observations in which the five behaviors occurred for each observation session in the *Hardlines* area of the store. Customer greeting, assistance offered, smiling, and small talk all remained at very low and stable levels during the initial baseline period. Eye contact fluctuated a bit more, but still remained at a consistently low level during the first baseline phase. There was a large

TABLE 2. Difference Scores, Pooled Standard Deviation, and Effect Sizes for All Dependent Variables Across All Phase Changes

Store Area/Behavior	FB-BL 1	BL 2-FB	Goals, FB, R+—BL2
HARDLINES			
Customer greeted mean difference	45.3	-35.5	42.1
<i>Pooled Standard Deviation</i>	22.09	21.45	23.33
<i>Effect Size (d)</i>	2.05	-1.66	1.8
Assistance offered mean difference	42.3	-31.7	41.5
<i>Pooled Standard Deviation</i>	17.68	18.18	23.92
<i>Effect Size (d)</i>	2.39	-1.74	1.73
Smiling mean difference	30.7	-20	35.3
<i>Pooled Standard Deviation</i>	14.28	13.09	19.57
<i>Effect Size (d)</i>	2.15	1.53	1.8
Eye contact mean difference	56.7	-48.1	49.9
<i>Pooled Standard Deviation</i>	25.38	25.3	26.42
<i>Effect Size (d)</i>	2.23	-1.9	1.89
Small talk mean difference	36.5	-31.6	34.7
<i>Pooled Standard Deviation</i>	15.07	17.8	19.05
<i>Effect Size (d)</i>	2.42	-1.78	1.82
SOFTLINES			
Customer greeted mean difference	54.8	-44.9	42.9
<i>Pooled Standard Deviation</i>	25.75	25.41	25.59
<i>Effect Size (d)</i>	2.13	-1.77	1.68
Assistance offered mean difference	51.8	-38.5	38.8
<i>Pooled Standard Deviation</i>	22.33	23.56	22.6
<i>Effect Size (d)</i>	2.32	1.63	1.72
Smiling mean difference	49.8	-41	44.6
<i>Pooled Standard Deviation</i>	21.48	23.77	24.35
<i>Effect Size (d)</i>	2.32	1.72	1.83
Eye contact mean difference	64.9	-49.1	45.2
<i>Pooled Standard Deviation</i>	27.61	25.93	24.74
<i>Effect Size (d)</i>	2.35	1.89	1.83

TABLE 2 (continued)

Store Area/Behavior	FB-BL 1	BL 2-FB	Goals, FB, R+ - BL2
CHECKOUT			
Customer greeted mean difference	56.2	-39.2	40.9
<i>Pooled Standard Deviation</i>	25.64	21.13	21.49
<i>Effect Size (d)</i>	2.19	-1.86	1.9
Smiling mean difference	50.8	-18.8	41.7
<i>Pooled Standard Deviation</i>	22.59	26.11	23.22
<i>Effect Size (d)</i>	2.25	-0.72	1.8
Eye contact mean difference	35.8	-32.1	32.1
<i>Pooled Standard Deviation</i>	22.59	17.24	17.15
<i>Effect Size (d)</i>	1.58	-1.86	1.87
Small talk mean difference	50.6	-48.9	48.1
<i>Pooled Standard Deviation</i>	21.51	25.31	25.88
<i>Effect Size (d)</i>	2.35	-1.93	1.86

increase for each of the five behaviors beginning in the first session of the feedback phase, and in general these increases occurred independently of the introduction of feedback across behaviors. However, the small increase in the percentages for smiling and eye contact during sessions 18 and 19 were likely due to some spread of the effects of feedback for customer greeting. Large increases in customer greeting resulted from presentation of the feedback for the four other dependent measures. The occurrence of each behavior decreased noticeably when feedback was withdrawn and remained at levels below those during the first intervention phase. There were increases in each of the five behaviors observed with the introduction of the second intervention phase. The average percentage of observations in which the behaviors occurred (see Table 1) was highest for all five behaviors during the second intervention phase.

Figure 2 depicts the percentage of observations in which each of the five customer service behaviors occurred per session for the *Softlines* area of the store. The graphs show a similar pattern of results as those found in the *Hardlines* area. Measures for all five behaviors remained low and steady throughout the first baseline phase. Performance was also higher for all five behaviors during both intervention phases than during either baseline phase. Increasing trends were observed during the

feedback phase for customer greeting, assistance offered, smiling, and eye contact, as well as for all behaviors observed during the goals, feedback, and celebration phase. Also, similar to the observations depicted in Figure 1, the increases in assistance offered, smiling, eye contact, and small talk during sessions 18 and 19 could have been due to some spread of effects of feedback for customer service. All five behaviors increased with the first session of both intervention phases, except for a one session delay of the increase in assistance offered during the feedback phase.

Figure 3 depicts the percentage of occurrences in which employees in the checkout lanes were observed greeting, smiling, making eye contact, and engaging in small talk with customers by session. There was more variability in the data during the first baseline phase for each measure in *Checkout* than in the other two areas of the store observed. The graphs show an increase in performance for the four behaviors during both intervention phases and an immediate decrease for each measure when the feedback was withdrawn. The increase in occurrences of eye contact during sessions 18 and 19 are likely a function of the increase in customer greeting during these sessions. Increases in customer greeting, smiling, and small talk during the first intervention phase appear to be independent of the introduction of feedback for customer greeting.

Figures 4, 5, and 6 show the final performance matrices for the three store areas. The matrices include the weekly goals and weights as determined by the store and department managers for each of the behaviors observed. The overall scores on each matrix equaled or exceeded the overall goal of 1000 in all three store areas for each of the three weeks the performance matrix was used in the store and most goals for specific behaviors were also achieved. The average scores for the three week period were 1000 for *Hardlines*, 1015 for *Softlines*, and 1020 for *Checkout*.

As described earlier, a second observer collected data during approximately 33% of the observation sessions to provide measures of interobserver reliability. Table 3 contains the interobserver reliability percentage scores by behavior and store area. Overall, there was a 92% agreement between observations made by the primary and secondary observer throughout the experiment. Interobserver reliability scores were lowest for eye contact, which averaged 89% and were highest for customer greeting which averaged 95%.

EFFECT SIZES

Table 2 presents data related to effect sizes for all dependent variables and phase changes throughout the study. For the first intervention

TABLE 3. Interobserver Agreement Results

Store Area/Behavior	Interobserver Agreement Result Percentages
HARDLINES	
Customer greeting	93
Assistance offered	91
Smiling	88
Eye contact	86
Small talk	90
SOFTLINES	
Customer greeting	93
Assistance offered	94
Smiling	91
Eye contact	87
Small talk	94
CHECKOUT	
Customer greeting	97
Smiling	95
Eye contact	92
Small talk	96

phase (feedback), effect sizes (d) in the study for each targeted behavior ranged from 1.58 to 2.42. For the Intervention Package II (Phase C) (feedback and reinforcement, goal setting, and celebration contingent on performance matrix scores), effect sizes ranged from 1.68 to 1.9. Effect sizes during this condition were smaller than those from Intervention I (Phase B) and this is most likely due to the fact that they were calculated using the difference between the second baseline and the second intervention phase. Because performance did not reverse to original baseline levels, the effect sizes therefore appeared to be smaller. In any case, according to Cohen's guidelines, an effect size of greater than .8 is normally considered "large." All of the effect sizes observed in this study were above 1.58 (Cohen, 1988).

Table 4 contains a summary of results regarding store leaders' delivery of feedback and praise to employees during the second intervention phase. It shows the number of times supervisors delivered praise while

TABLE 4. Number of Times Supervisors Delivered Praise on Store Floor to Employees During Phase C by Store Area (from Reinforcement Logs)

Store Area/ Behavior	Number of Times Supervisors Delivered Praise
HARDLINES	
Customer greeting	52
Assistance offered	40
Smiling	47
Eye contact	52
Small talk	8
SOFTLINES	
Customer greeting	61
Assistance offered	45
Smiling	38
Eye contact	55
Small talk	12
CHECKOUT	
Customer greeting	38
Smiling	36
Eye contact	38
Small talk	20

on the store floor to employees, based on self-reported data from the managers’ reinforcement logs. A sample reinforcement log data sheet is shown in Figure 7. Supervisors or managers provided employees with feedback at team meetings during each of the three weekly meetings of the second intervention phase, for each area of the store, respectively.

DISCUSSION

This study was an attempt to positively impact a variety of customer service behaviors for the majority of employees working storewide in a large department store setting. Based on the data summarized in Figures 1, 2, and 3, it is clear that the presence of both the feedback intervention and the package intervention (incorporating feedback and reinforce-

FIGURE 7. Reinforcement Log Data Sheet

REINFORCEMENT LOG			
Observer: _____	Store Area (circle):		
Date: _____	Hardlines	Softlines	Checkout
Time: _____			
	Positive Feedback Given to Employee?		
	YES?	NO?	
Customer Greeting:	_____	_____	
Assistance Offered:	_____	_____	
Smiling:	_____	_____	
Eye Contact:	_____	_____	
Small Talk:	_____	_____	

ment, goal-setting, and a celebration contingent on the measures posted on the performance matrix), substantially increased all five customer service-related behaviors for employees working on the sales floor (i.e., *Hardlines* and *Softlines*) and also the four customer service behaviors for employees working in *Checkout*. In general, the effects of implementing and removing the first intervention were immediate and large, ranging from 1.58 to 2.42 (see Table 2 for a summary of all effect sizes). Therefore, there appears to be good evidence that the interventions caused the increase in the occurrence of the behaviors observed, and this conclusion is particularly strengthened by the consistent results seen in the reversal (for the first intervention) and AC treatment effects (for the second intervention package), across the three store areas and 14 behaviors.

There were differences, however, in the levels of performance across the three store areas observed. The four behaviors measured at *Checkout* were at higher levels than the corresponding measures in the *Hardlines* and *Softlines* areas throughout the study. This can be explained by at least several factors. Employees working at the checkout lanes generally didn't engage in the same number of service-incompatible behaviors or have the same amount of reinforcement available to them for such competing behavior as did employees working on the sales floor. Employees working on the sales floor were often required to stock shelves and to set up displays and these behaviors may have occasionally interfered with optimal customer service behavior. Also, em-

ployees working on the sales floor were often observed walking and talking with coworkers and/or talking with coworkers on walkie-talkies. Although employees working at the checkout lanes occasionally were observed talking to coworkers while helping customers, the number of opportunities to socialize with coworkers were many fewer in the checkout lanes than they were on the sales floor. Further, employees on the general sales floor could often avoid interacting with a customer by walking away from the customer. This was not possible for employees at the checkout lanes. There also might have been a greater response effort associated with approaching customers on the sales floor than for greeting and talking to customers who were already present at a checkout lane. This assumption was perhaps supported by the fact that during a meeting with a manager of *Checkout*, the manager reacted with disappointment when shown the graphs of baseline performance. His reply was that it shouldn't be "too much to expect" that employees would greet almost every customer in a friendly manner as they checkout given the low response effort required.

There is at least one possible explanation that may account for the higher levels of performance for all five behaviors observed in *Hardlines* and for all but two of the behaviors observed in *Softlines* and *Checkout* during the second baseline phase. Since performance during the first baseline period was especially low in the *Hardlines* and *Softlines* for most behaviors measured (and was often at 0%), many employees may have only infrequently encountered positive reactions from customers, and employees may have never (or rarely) experienced positive or negative reactions from managers regarding customer service. Behavior may have remained low during the first baseline partially for these reasons. Performance increased markedly during the first intervention phase, and many employees were initiating contact with customers at a much higher rate than during the first baseline phase. Although no formal data were collected on customer behavior, anecdotally speaking, there was noticeably more laughing, smiling, and small talk from customers during the intervention phase. These reactions might have functioned as natural reinforcement for employees and therefore, might have served to alter many employees' behavioral repertoire for the target behaviors long enough to affect performance measures during the second baseline phase.

Unfortunately, the intervention was not continued at the store following the end of the data collection period for the study. This was due to various factors. First, the store changed store managers twice since termination of the study. The commitment of the later store managers was

never as high as the first store manager who championed the intervention. Second, the store manager at the end of the study, although very pleased with the results, felt the results could be maintained without further data collection. Since the store and department managers were taught to recognize and reinforce the observed behavior on the floor and filled out reinforcement logs, it seems likely that they could have, without too much effort, continued to collect data after the experiment ended. Although the store manager opted to discontinue data collection, the store and department managers reported that employees as a whole liked the store celebration, which was awarded for all three store areas, and that employees appreciated getting feedback as a group. When asked why employees liked getting feedback one department manager said that employees liked it because the data “were always used in a positive way.”

It is clear that customer service behaviors, due to their intangible nature, can be difficult to manage. Results from customer satisfaction surveys provide the only programmed consequence many employees receive for customer service related behaviors, and due to the cost and difficulty of administration, customer satisfaction surveys are rarely collected frequently enough to provide meaningful feedback to employees. Therefore, the only immediate consequences employees receive for engaging in courteous behavior are naturally occurring ones (i.e., directly from the customer). Unfortunately, the types of consequences employees receive from customers vary depending on the social skills of both the employee and the customer. Sometimes excellent customer service behavior is met with extinction or even social punishment from customers. The behavioral effects of such consequences are to lower the likelihood of the employee emitting those customer service behaviors in a similar situation in the future. Customer service behavior may be further weakened by the poor social skills of the performer, and this can reduce the probability that the customer will react in a positive manner to the interaction. An idea that has yet to be investigated in the area of customer service is to evaluate the impact of customer-driven interventions. Rohn, Austin, and Sanford (2003) evaluated such a strategy in a fast food establishment by having *customers* deliver prompts, feedback, and praise to employees, in order to improve *employee* suggestive selling. In the absence of such regular customer-driven consequences, to achieve and maintain high levels of customer service from all employees, regular and direct observation, frequent performance feedback, and reinforcement may be required. The current study systematically made use of the latter strategy.

Since individual performance data were not collected during the study, the relative effects of the goals, feedback, and reinforcement on particular individuals could not be assessed. It may have been that training involving practicing customer service interactions could have facilitated improvements in courtesy. Brown et al. (1980) used a training program that required participants to practice the customer service behaviors being measured. The training program seemed to produce slight increases in performance, although a feedback intervention was even more successful. However, since only three participants were involved in that study, the impact of customer service training involving behavioral practice on a larger scale would still be worthy of investigation.

The current study has many limitations, however, we focus on only the three discussed below. The experimental design represents a primary limitation in that only one treatment effect (the feedback phase, or B phase) was demonstrated through reversal within each baseline. The graphic feedback treatment (phase B) was effectively evaluated, given there was a treatment effect (with 14 cross-baseline replications) and a reversal during the second baseline (with 14 cross-baseline replications of the reversal). However, it should be noted that the effects of the second treatment (the package intervention, or phase C) were not demonstrated as strongly with respect to the experimental design used in the study. That is, given that the design was an ABAC, the second treatment phase had no reversal phase to confirm its efficacy. However, the second intervention phase did have 14 cross-baseline replications, representing a considerable improvement compared to a simple AB design. Still, because the replications occurred so close in time to one another, maturation is a factor that might have contributed to the efficacy of the second treatment package. Also, the rather quick reduction in the occurrence of the behaviors observed during the reversal suggests the possibility that some other type of communication by managers to employees occurred in addition to the instruction that they were only to tell employees that the graphs were no longer going to be posted. At the same time, there is no obvious source of potential reinforcement for managers to engage in this sort of behavior, assuming that their store manager remained interested in supporting improved quality service among all employees.

Another limitation of the study was that the individual differences in reactions to the interventions among employees observed during the study were not collected. Since the study included observations of a large number of employees (115) and because observations were obtained without the employees being aware when they were being ob-

served, it would have been difficult to collect this type of data. However, it would have been valuable to have data reflecting how many different employees were being observed and also data that showed how customer service might have differed across employees with different levels of job experience and other individual differences that might be related to service behavior.

Also, although the store manager indicated that there was little turnover and no atypical hiring or workforce reduction during the period in which observations were conducted, specific data on turnover were not collected. These data would have been useful since turnover can be a potentially confounding variable. However, it seems unlikely that turnover had much effect on the changes observed in the behaviors observed because there was no unusual change in workforce composition that occurred across the three store areas at the time the interventions were introduced and the reversals occurred.

Finally, the study shares an almost ubiquitous problem among field experiments in the OBM tradition. That problem concerns the difficulty of gaining access to research sites for which managers or administrators will permit researchers to examine a limited number of intervention components for the sake of isolating effects of specific components because organizational authorities, rightfully, insist on assurances of desired results rather than accept payoffs denominated in knowledge regarding relative effectiveness of intervention components. In the present study, for example, it would have been preferable to have compared the effects of feedback in the first intervention phase with effects of goals and feedback as points earned using the matrix alone, i.e., without store area celebrations and increased managerial administration of social reinforcements. Should an opportunity to do so present itself to researchers in the future, therefore, they should attempt to compare effects of a performance matrix with and without other intervention components known to contribute to employee performance improvements.

Although this version of the performance matrix tool has been discussed and reportedly used in practice for more than 14 years since Daniels (1989) described it, this study was the only one our search located that reported results from the use of the performance matrix (as described by Daniels—see our discussion in the opening pages of this manuscript for some notable exceptions) to help change employee behavior in an organizational setting. The pros and cons associated with using a performance matrix in real organizations should be examined in more detail in future studies. Although it wasn't observed to be the case

in this study, in some instances understanding how a performance matrix works may be unnecessarily confusing to some managers or employees. Research designed to examine the difficulties individuals may experience in understanding the matrix, and if the matrix is an effective treatment apart from other package elements may be important. Also, research aimed at establishing the generality of the performance matrix's effectiveness when work behaviors other than customer service are targeted could be an important addition to the literature.

REFERENCES

- Abernathy, W. B. (2001). An analysis of the results and structure of twelve organizations' performance scorecard and incentive pay systems. In L. Hayes, J. Austin, R. Houmanfar, & M. Clayton (Eds.) *Organizational change* (pp. 240-272). Reno, NV: Context Press.
- Albrecht, K., & Zemke, R. (1985). *Service America!* New York: Warner Books, Inc.
- Anonymous (2002, November). *Kohl's ranks highest in customer satisfaction among moderate-price department stores Wal-Mart ranks highest in customer satisfaction among discount department stores*. Retrieved December 8, 2004, at http://www.jdpa.com/studies_jdpower/pressrelease.asp?StudyID=687
- Austin, J. (2000). Performance analysis and performance diagnostics. In J. Austin & J.E. Carr (Eds.), *Handbook of applied behavior analysis* (pp. 321-349). Reno, NV: Context Press.
- Austin, J., Carr, J. E., & Agnew, J. (1999). The need for assessing maintaining variables in OBM. *Journal of Organizational Behavior Management*, 19(2), 59-87.
- Austin, J., Weatherly, N., & Gravina, N. (in press). Using task clarification, graphic feedback, and verbal feedback to increase closing task completion in a privately owned restaurant. *Journal of Applied Behavior Analysis*.
- Baum, W. M. (1994). *Understanding behaviorism: Science, behavior, and culture*. New York: HarperCollins.
- Brown, C.S., & Sulzer-Azaroff, B. (1994). An assessment of the relationship between customer satisfaction and service friendliness. *Journal of Organizational Behavior Management*, 14(2), 55-75.
- Brown, M.G., Malott, R.W., Dillon, M.J., & Keeps, E.J. (1980). Improving customer service in a large department store through the use of training and feedback. *Journal of Organizational Behavior Management*, 2(4), 251-265.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. New York: Academic Press.
- Crowell, C.R., Anderson, D.C., Abel, D.M., & Sergio, J.P. (1988). Task clarification, performance feedback, and social praise: Procedures for improving the customer service of bank tellers. *Journal of Applied Behavior Analysis*, 21(1), 65-71.
- Daniels, A.C. (1989). *Performance management: Improving quality productivity through positive reinforcement*. (3rd ed.). Tucker, GA: Performance Management Publications.

- Gilbert, T.F. (1978). *Human competence: Engineering worthy performance*. Washington, DC: International Society for Performance Improvement.
- How to measure customer satisfaction*. (2002). Retrieved December 8, 2004 from <http://www.woodsco.com/CustomerSatisfaction/Customer%20Satisfaction.htm>
- Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard: Measures that drive performance. *Harvard Business Review*, 70(1), 71-80.
- Komaki, J., Blood, M.R., & Holder, D. (1980). Fostering friendliness in a fast food franchise. *Journal of Organizational Behavior Management*, 2(3), 151-164.
- Luthans, F., Paul, R., & Taylor, L. (1986). The impact of contingent reinforcement on retail salespersons' performance behaviors: A replicated field experiment. *Journal of Organizational Behavior Management*, 7(1/2), 25-35.
- Malott, R. W. (1992). A theory of rule-governed behavior and organizational behavior management. *Journal of Organizational Behavior Management*, 12(2), 45-65.
- Rachlin, H. (1989). *Decision, decision, and choice*. New York: W. H. Freeman.
- Riggs, J.L. (1986). Monitoring with a matrix that motivates. *Performance Management Magazine*, 4(3), 13-19.
- Rohn, D., Austin, J., & Sanford, A. (2003). A consumer-driven approach to increasing suggestive selling. *Performance Improvement Quarterly*, 16(1), 29-39.
- Stoneman, K.G., & Dickinson, A.M. (1989). Individual performance as a function of group contingencies and group size. *Journal of Organizational Behavior Management*, 10, 131-150.
- The Age of Wal-Mart: Inside America's most powerful company*. (2004, November 10). [Television broadcast]. Transcript available at: <http://cnbc.burrelles.com/product.asp?ProductID=439911>
- Welsh, D.H.B., Bernstein, D.J., & Luthans, F. (1992). Application of the Premack principle of reinforcement to the quality performance of service employees. *Journal of Organizational Behavior Management*, 13(1), 9-32.
- Wilk, L.A., & Redmon, W.K. (1990). A daily-adjusted goal setting and feedback procedure for improving productivity in a university admissions department. *Journal of Organizational Behavior Management*, 11, 55-75.