Behavioral System Feedback Measurement Failure: Sweeping Quality Under the Rug

Maria T. Mihalic and Timothy D. Ludwig

ABSTRACT

Behavioral Systems rely on valid measurement systems to manage processes and feedback and to deliver contingencies. An examination of measurement system components designed to track customer service quality of furniture delivery drivers revealed the measurement system failed to capture information it was designed to measure. A reason for this failure was an inadequate design, which resulted in sabotage of the measurement system by agents in the organization. The failure of this measurement system led to the ineffective operation of business processes and related systems. For example, a bonus pay system and a disincentive system were disabled due to the faulty measurement system. Suggestions for the development of a valid measurement system are offered.

ARTICLE

Behavioral Systems Analysis (BSA) involves analyzing components of interconnected processes that impact individual and organizational performance (Brethower, 1982, 2000, 2001, 2002; M. E. Malott, 2003; Sulzer-Azaroff, 2000). According to Malott (2003), individuals behave in ways that should be consistent with the goals and mission of their organization. Individuals working together toward goals do so in organized processes that include the sequenced behaviors of individuals (or machines) across different functions and levels of the organization (Rummler, 2004). Brethower (2000) argues that these processes must be monitored via process and customer

feedback to better align the processes with the system goals.

Measurement is a key component of a healthy behavioral system. Techniques used in BSA require effective monitoring through measurement systems focusing on critical steps across organizational levels (Abernathy, 2000; Glenn, 1988; R. W. Malott, 1974, 1999; Rummler, 2004). When process data are collected and analyzed they can then provide feedback to individuals responsible for the process (Brethower, 2000). Process feedback can be used to shape behaviors necessary to accomplish the organization's mission (M. E. Malott, 2003). If process feedback is not available, managers of the process have little basis for improving the processes and the behavioral contingencies therein. Processes become neglected (Katz & Kahn, 1966), and other organizational systems also fail to function. Thus, measurement systems provide the process feedback necessary for system survival. Measurement systems are used to assess employees, work groups, departments, or entire organizations with a variety of organizational measures (e.g., absenteeism, productivity, quality, etc.) that directly relate to critical business issues (Rummler, 2004). Measurement systems are designed not only to assess current levels of performance, but also to provide feedback to shape performance toward goals, aid in organizational diagnosis and improvement, and provide the foundation for bonuses and promotions (Anderson & Fagerhaug, 2002). "Measurement is the first step that leads to control and eventually improvement. If you can't measure something, you can't understand it. If you can't understand it, you can't control it. If you can't control it, you can't improve it" (Harrington & McNellis, 2006).

Measurement Systems

In the measurement literature, this feedback in behavioral systems has been referred to as "organizational control systems." Organizational control is the "process of influencing members of a formal organization" (Flamholtz, 1979, p. 51). Organizational control systems are devised to increase the likelihood that members of an organization behave in ways that are consistent with organizational goals through the delivery of contingent performance feedback (see also M. E. Malott, 2003). Control systems allow leaders to regulate the activities of the organization and monitor achievement toward performance standards (Lewis, Goodman, & Fandt, 2004). Power (2004) explained that without effective measurement systems, organizations would be hindered by individuals who act in their own interests instead of those who help achieve organizational goals.

Lewis and others (2004) outline the four major components of an effective organizational control system. The first component of a control system is the establishment of organizational goals. Hall (1975) explained that goals are important because resources can be distributed and policies can be created based on them. The next component is the establishment of standards of performance. In the next component, measurement is used to evaluate performance against these standards to discover gaps. Lastly, rewards can be assigned for surpassing the standard of performance, while corrective actions address deviations from the goal that are not acceptable.

Measurement systems provide a framework from which decisions can be made and resources allocated (Anderson & Fagerhaug, 2002; Kula, 1986). Data produced by measurement systems must allow the organization to predict future performance (Bassi & McMurrer, 2005) and serve as the foundation of many business decisions such as daily operations, planning, and early warning systems (Anderson & Fagerhaug, 2002).

To eliminate situations where no contingencies exist and the behavior goes unchanged, measurement systems serve as the basis for incentive plans designed to reinforce behavior (Anderson & Fagerhaug, 2002). However, incentives can sometimes be fatal remedies when based on faulty measurement systems that create incentives for behaviors counter to the actual goals of the organization (Sieber, 1981).

Components of a Good Measurement System

To identify gaps in performance, Bassi and McMurrer (2005) state measurement systems should be "actionable," focusing only on items that are under the control of the organization; "descriptive," summarizing key system issues through statistics; and "detailed" to enable the organization to pinpoint behaviors in need of attention. A measurement system also needs to be accurate and consistent, producing unbiased, reliable data. Reliability is the degree of stability and consistency inherent in a set of data (Bernstein, Penner, Clarke-Stewart, & Roy, 2003). "Measurements are *in principle* replicable and not dependent on when, where, and by whom the measurement is done" (Power, 2004, p. 769).

A measurement system should be parsimonious enough to facilitate reliable application and results. As a measurement system becomes more complicated, the reliability of the system tends to be challenged. An increase in variables to measure leads to an increase in measurement error (such as in code entry), which reduces reliability. In turn, user friendliness will increase the reliability of the measurement system.

Bernstein and others (2003) explained that the usefulness of a measurement system is represented by the validity of the data produced. A number of factors can influence validity. The content validity of a measure is the degree to which the measure relates to the variable being assessed (Bernstein et al., 2003). Furthermore, a measure with high content validity samples all the variables related to the topic, not a small segment (Lanyon & Goodstein, 1997). An invalid measure may not sample enough aspects of the variable or may assess components not related to the variable.

A measurement system must facilitate timely delivery of feedback to employees to facilitate change and improvement in the organization (Anderson & Fagerhaug, 2002). The more proximal feedback delivery is to performance, the more effective the feedback is (Ludwig & Goomas, 2007). Feedback delivered proximal to performance allows the performer to gain access to information close to the emitting of behavior and more immediately

correct errors. As time passes between behavior and feedback, the link between the behavior and its consequence is weakened, and feedback becomes less effective (Daniels & Daniels, 2004).

Finally, Garnego, Biazzo, and Bitici (2005) state that measurement systems should include a review function that allows the system to be adapted to respond to changes and to fit with the strategy of an organization. A static measurement system in a dynamic organizational environment will quickly become obsolete. To ensure continued relevance, a measurement system must be dynamic and allow for change when needed. The system should be designed to allow employees and users to provide input to improve the utility of the measures and systems (Anderson & Fagerhaug, 2002).

In summary, an effective measurement system must, first and foremost, be valid and reliable. In addition, the system needs to be dynamic enough to change with the organization. It must provide accurate and understandable feedback to employees and managers to monitor and help change behavior. Organizations that follow these guidelines measure meaningful variables accurately and are in a better position to manage control over organizational goals.

Incentive Systems

Incentive systems have been shown to decrease the amount of monitoring necessary for supervisors to maintain performance (Banker, Lee, Potter, & Srinivasan, 1996) if made contingent on performance (Bucklin & Dickinson, 2001). Pritchard, Jones, Roth, Stuebing, & Ekeberg (1988, 1989) found that the addition of feedback to measurement systems increased the productivity of employees 50% over baseline levels. Furthermore, adding incentives increased productivity 76% over baseline performance.
Incentives also serve as feedback for the performer (Goomas & Ludwig, 2007). Indeed, the effects of feedback can be strengthened with the addition of rewards (Alvero, Bucklin, & Austin, 2001; Balcazar, Hopkins, & Suarez, 1985/86; Welsch, Luthans, & Sommers, 1993). Thus, incentive systems have been found to be effective in organizations when contingent on desired behavior and paired with frequent feedback (Abernathy, Duffy, & O'Brien, 1982; Bateman & Ludwig, 2003; Dickinson & Gillette, 1993; Gaetani, Hoxeng, & Austin, 1985; Honeywell-Johnson, McGee, Culig, & Dickinson, 2002). However, if the measurements that serve as the basis of the incentive system are flawed, agents of the system may behave in ways to gain the incentive (or avoid the disincentive) that may not support the goals of the system.

CASE STUDY

The present report looked at the measurement system of a high-end furniture distribution company to identify weaknesses in design and implementation, and to interpret the causes of these weaknesses in context of a number of theories for management, economic, and psychological literature including

organizational control, agency theory, and behavioral analysis.

Company Profile

This study was conducted at a furniture distribution company located in the southeastern United States that catered to high-end clientele throughout the country. This company received, stored, and distributed approximately 60,000 furniture orders per year with an average of approximately $160 million in delivered sales yearly.

Customer orders were received, and orders were placed with the manufacturer. Furniture pieces were then received from the manufacturer and stored in the warehouse. When all the pieces of a customer order were received by the warehouse, items were assembled, inspected, and repaired (if necessary). After the orders were prepared, furniture was loaded on a truck and delivered to a customer's house, office, store, or other location.

There were numerous errors that could occur during the delivery process, such as transit damages and leaving parts in the truck. When damaged or imperfect furniture was discovered after arriving at the customer location, drivers were required to call the company's Customer Service Department for authorization to fix damaged pieces, schedule an outside repair, or return the furniture to the truck. Customer Service analyzed the estimated cost of the repair performed on site versus the cost of bringing the piece back to the warehouse for repair and redelivery before giving authorization. However, drivers employed by this company were leaving damaged, imperfect, and even incorrect furniture in the customers' houses without calling Customer Service.

Left in Home Unauthorized

A Left in Home (LIH) error was defined as any piece of furniture in need of repair or return left in the delivery location without calling a customer service representative (CSR) for authorization. LIH errors also occurred when drivers waited to call CSR until after leaving the customer home or when the customers themselves called about a damaged or incorrect piece left in the home. Finally, the driver may report the LIH when the truck returned to the warehouse at the inbound docks. In each case, the damaged or incorrect piece was left in the delivery location and CSR was not called while the driver was at the customer location, therefore the driver did not receive authorization to leave the piece(s). LIH incidents resulted in the company having to contract a local professional to do the repairs at higher costs or send a truck out to pick up the furniture piece for repair and redelivery.

The company had developed an elaborate listing of error codes for their measurement database to aid in the tracking and resolution of errors in delivery. The error code, description of the problem, furniture piece identification number, and actions taken were put into the database by either a

customer service representative or the manager of the inbound delivery docks. The identifying number of the furniture piece could be tracked to the customer order and delivery team.

Driver Pay System

In addition to tracking errors, the measurement system was designed to input performance data into a pay system for delivery drivers. On a biweekly basis, drivers received a "load pay" amounting to 6.2% of the revenue of each order delivered to the customer. The driver team lost the load pay for a delivery if an error was recorded. An additional 25% of the load pay was eligible for a quarterly bonus. If errors were extensive, 10% of the bonus was forfeited. Moreover, an immediate charge of $200 was assigned automatically when LIH error codes were assigned.

Major Agents of the Measurement System

In addition to the CSRs and inbound managers who entered error codes, a number of other agents had an impact on the measurement system or were subject to contingencies based on the measurement system.

DRIVER

Delivery drivers delivered customer orders, unloaded and set up the furniture in the customer home or office, and inspected the furniture for defects. If any problems with the furniture piece were discovered during installation, drivers were trained to repair the piece. If the repair could not be done, the driver was to call Customer Service to determine whether to contract a local repair person or have the piece returned to the warehouse. After all orders were delivered, drivers returned to the warehouse, turned in paperwork, and unloaded any furniture items left on the truck.

CUSTOMER SERVICE REPRESENTATIVE

If any problems arose during delivery, a CSR became the point of contact for resolving customer issues and recording errors in the Customer Service database. The CSR opened a problem slip in the database and recorded relevant customer and delivery information before describing the specific error in detail. The CSR then applied the most representative error code out of the possible 145 error codes in the Customer Service database. CSRs added information to problem slips to document steps made to fix the problem (e.g., repairs, subsequent delivery).

INBOUND MANAGER

When driver teams returned to the warehouse they were asked to recount

any delivery errors that may have occurred. This information was recorded by the Inbound manager in the Inbound database with accompanying delivery and customer information. Error codes were applied from the list of 99 Inbound problem codes.

The Inbound database used a different coding system than the Customer Service database. The Customer Service database was the main measurement system designed to track errors and provide information for the company's bonus pay systems. The Inbound measurement system, on the other hand, was designed to track problem furniture pieces as they moved through the resolution process.

AUDITOR

To check for accuracy, auditors monitored the application of codes in the Customer Service database. The audit procedure required that full records of errors affecting driver pay be reviewed. Written information in this database was compared to the CSR error code list to ensure the error code applied was accurate prior to bonus pay periods.

QUALITY SERVICE MANAGER

Quality service managers combined the Customer Service database and the Inbound database. The data from each database were extracted, combined, and aggregated to generate reports for various levels of management. These reports summarized the number of each error code monthly and were used to monitor error rates across variables of interest (e.g., driver, furniture brand, and store). Quality service managers also monitored both databases for irregularities and inconsistencies in reporting.

VICE PRESIDENT OF OPERATIONS

Among other responsibilities, the vice president of operations was in charge of monitoring and managing the performance of the drivers. Thus, the measurement and bonus pay systems were under his direct control. Reports generated by this measurement system were used to make his managerial decisions regarding personnel, customer service, and financial investments or cost cutting. It was the vice president who initially developed the disincentive system designed to reduce driver errors, including LIH.

INVESTIGATION

An investigation of the company's measurement system data was undertaken to further pinpoint the variables of interest. This investigation specifically centered on the Customer Service database as the primary source used

by the pay systems to evaluate driver performance and to dispense bonuses and penalties. This database housed the problem slips opened by the CSRs to document errors.

Data were evaluated to examine error trends over a 7-year period, from 2000–2006. During 2000 and 2001 the number of LIH errors went from 4,732 to only 44. The number of authorizations (where drivers got permission to leave defective pieces in homes) went from 0 in 2000 to 3,235 in 2001. Very low levels of LIH were maintained through 2006 (see Figure 1). A preliminary
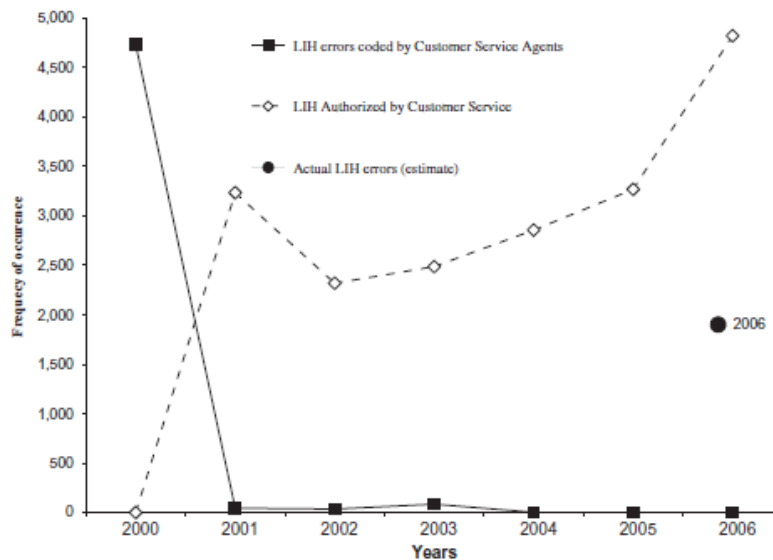

FIGURE 1 LIH Errors: 2000–2006. Filled squares represent annual LIH errors coded by CSRs. Open diamonds represent annual authorizations of LIH by CSRs. The filled circle represent an estimate of actual LIH errors for 2006 based on our quality service audit.

hypothesis was that the substantial decrease in LIH errors was due to the implementation of the two disincentives (i.e., $200 charge in load pay and forfeiture of bonus pay) within the bonus pay system in October 2001. The initiation of the disincentive system was credited by the vice president to help alleviate an expensive problem for the company. It is also noteworthy that, from its inception in 2001, the measurement system seldom resulted in applying any disincentives to drivers for LIH errors.

After further analysis it became apparent that the decrease in the number of LIH errors could not have been due to the disincentive. The disincentive was put in place in October of 2001; however, the decrease in the number of LIH errors seemed to begin in the first week of 2001. The question then became, "What happened at the beginning of 2001 that could account for such a drastic decrease of errors?"

Interviews with management revealed that a new measurement system had been put in place at the beginning of 2001 to more accurately capture, report, and control errors. Year 2000 errors were not recorded by CSRs. Instead, reports were generated by reading the original customer service records, and errors for the year 2000 were recoded according to the new measurement coding system implemented in 2001. This post-hoc coding resulted in the wide discrepancy in the frequency of LIH errors between the years 2000 and 2001 (see Figure 1).

Additionally, there was a consensus among employees consulting with the project that LIH errors were not being accurately captured. Interviews with various staff members revealed that employees had learned to manipulate the measurement system. Thus, a more thorough investigation was undertaken to reevaluate the data.

For this investigation the quality service manager and another colleague analyzed Customer Service problem slips by hand for the month of August 2006, in the same way the year 2000 errors were coded previously. These paper records contained problem descriptions written by CSRs, including error codes assigned and details of the problems entered into a free response section. The details of the problem often revealed other errors that occurred during a delivery that were not assigned an error code. For example, an error code for "transit damage" may have been filed, but the written information stated that the furniture piece was left in the home without authorization. Thus, the error should also have been coded as LIH.

An analysis of 160 problem slips from August 2006 determined that 63 incidents (39.4% of the 160 incidents) should have been coded as LIH but were not. The Customer Service measurement system for this time period reported only 2 occurrences of LIH. Assuming that approximately 40% of the problem slips should have been recorded with an LIH code, as many as 1,901 incidents of 4,824 problem slips opened in all of 2006 should have been recorded as LIH (see Figure 1). This estimated discrepancy suggested the system was not capturing LIH errors, and the measurement system had failed.

FAILURE OF THE MEASUREMENT SYSTEM

Too Many Error Codes

Tangen (2004) explained that information overload can occur from too many performance measures, which results in inaccurate recording. As a result of the over 200 different codes, CSRs were applying only a handful of codes, mostly the general problem codes that they were most familiar with. CSRs reported in interviews that many of the codes were considered redundant, did not describe good performance, and were never used for reporting. The abundance of error codes greatly reduced the reliability of the data because different CSRs picked different error codes given the same information. No real operational definitions of these codes existed for the CSRs. Even employees who developed the system had differing opinions during interviews about what constituted various errors.

To complicate matters, new codes could easily be created in the system with no prior approval. The system was designed to describe the error from the broadest category (i.e., what the error was) to the most precise (i.e., who did it, what furniture pieces were damaged, etc). However, the system had not maintained this order, nor was it ever clearly established according to this goal. Yet the system administrators and executives in the organization refused to simplify the system and reduce the number of codes used, even though it was common knowledge that fewer than 10 codes were being applied routinely. After all, "it is a waste of time to collect data if they are being ignored" (Tangen, 2004, p. 728).

Inaccurate Feedback

The company's measurement system, designed to provide feedback for correcting gaps in performance, could not accurately or consistently identify performance errors. The reduced reliability and validity of the data resulted in inaccurate feedback. As a result, the employees monitored under this measurement system were incorrectly being told that they were behaving correctly. Similarly, reports using the system misled management by suggesting that customers were being left with high-quality furniture nearly 100% of the time.

Misapplied Incentive Systems

The failure of the measurement system led to a failure in the incentive system. Latham and Dossett (1978) explained that failure of incentive systems can be linked to money not being made contingent on verified performance. Additionally, the lag time between when data were entered and when drivers got feedback was often months. Drivers were supposed to be notified of errors via biweekly paychecks, when disincentives related to specific error codes were withheld from their pay. The majority of the time, however, drivers were made aware of errors only in an aggregated report in quarterly bonus. This lag time and the aggregation of target errors made it improbable if not impossible for drivers to link specific instances of errors to their feedback and incentive pay.

Undermining the Measurement System

Our investigation revealed that the measurement system seriously underreported thousands of LIH errors that may have been occurring each year. Thus, the incentive system may have ended up reinforcing behaviors that manipulated the system to avoid error code assignments. Indeed, drivers interviewed said they learned to call CSRs after multiple deliveries to avoid the LIH code or to wait to report errors to Inbound. However, weaknesses in accountability were not limited to drivers. There were also weaknesses in the recording, checking, and reporting of error codes.

CORRUPTION OF THE MEASUREMENT SYSTEM

Agency theory (Jensen & Meckling, 1976) explores why the agents of the system begin acting in opposition to the system's goals. In concert with behavior analysis (Skinner, 1953), we can examine specific contingencies that may have encouraged the manipulation of the measurement system for each major agent in the organization.

Agency Theory

Agency theory describes problems that arise in an organization by considering what occurs when different goals exist between cooperating parties (Eisenhardt, 1989). Agency theory has been applied to economics (e.g., Jensen & Meckling, 1976), sociology (e.g., White, 1985), and organizational behavior (e.g., Eisenhardt, 1985) and has resulted in numerous literature reviews (Eisenhardt, 1989; Mitnick, 1992, 1998; Moe, 1984).

Much of the research in this area revolves around the assumption that agents are "pursuing self-interest with guile" (Williamson, 1975, p. 134). Individuals act in a self-serving manner instead of acting in the best interest of the company, often to protect themselves from risk (Shapiro, 2005). All of the major agents in this case study had a number of reasons to sabotage the system in order to protect their own interests or the interests of others.

According to Shapiro (2005), one of the problems that arise is what he calls "information asymmetries" and centers on the organization's inability to know what its agents, or employees, are doing. This is often because the information cannot be obtained without great difficulty or expense (Eisenhardt, 1989). The failure, in this case, of a measurement system to adequately inform the organization of its agents' activities resulted in unreported errors in customer service. The agents of the measurement system in this case study began acting in opposition to the goals of the system.

Behavior Analysis

Behavior analysis examines the relationship between behavior and the environmental consequences of behaviors (Rogers & Skinner, 1956; Skinner, 1953). An important consideration is the response cost of the desired versus actual behaviors. Response cost is the effort required to do the behavior or to obtain a result based on the behavior(s) (Gormezano & Wasserman, 1992). The more effort required to achieve a result, the less likely the agent is to perform these behaviors. Response cost can punish responding when the consequence, or cost, of performing is increased effort on the part of the agent. In the present case, error reporting was laborious and, as a result, LIH data were not collected. In the context of the behavioral system, each agent had a variety of contingencies that could easily be used to explain their resistance to using the system as it was designed. Descriptions of the contingencies on each agent are detailed below.

DRIVERS

While delivery drivers were to call in and report LIHs, there was almost no reinforcement for doing so. Instead, if the drivers called in to report an error, they lost the pay they earned for that delivery. The load pay for the trip was paid to the driver teams only if the delivery was carried out without any error. Thus drivers began calling in to report the error after leaving the delivery location or reporting the error at Inbound to avoid the disincentive. In addition, drivers experienced response costs by making phone calls to Customer Service and going through the process of obtaining approval. By not calling from the customer's home, drivers could avoid having to move the furniture piece back into the truck, reshuffling the rest of the day's deliveries to make room, and unloading it at the warehouse.

CUSTOMER SERVICE

The correct application of the error codes by CSRs would have led to accurate records and reports, which may have resulted in praise from supervisors. However, these consequences were not salient to CSRs because reinforcement resulting from accuracy may never have occurred. In contrast, applying multiple error codes incurred response costs. So instead of opening separate problem slips for each separate error that occurred during a delivery, the CSR tended to report only one error and document the details of the delivery error in the trip detail section of the single problem slip. The LIH error did not get coded, and the text describing it became buried in the paperwork, which was not easily searchable and did not inform any of the measurement or pay systems.

Further, incorrect error coding allowed CSRs to avoid uncomfortable social situations. The CSRs were being asked to help in administering the $200+ punishment to drivers. CSRs reported that the male drivers would boisterously cajole the female CSRs to change the code after the fact. Thus, recording errors resulted in arguments with drivers, which socially punished CSRs for applying the correct codes. Thus, CSRs learned over time that incorrectly coding these errors saved time, effort, and social unpleasantness.

AUDITORS

To verify the validity of the data produced by the measurement system, the organization appointed auditors to review Customer Service records for inaccurate error codes before issuing quarterly bonuses. This entailed the review of thousands of records for the quarter, which auditors reported as "tedious and boring." In order to avoid the response costs auditors would procrastinate. According to interviews, auditors tended to wait until the last minute to perform the audits because these duties interfered with regular tasks required in their job. In practice, auditors put off the task for so long others had to review the records for them. Sometimes extensions were applied for and/or the review was never completed. Thus, there were not sufficient reinforcers to offset the response costs of correctly doing the

audits, and this eliminated the only source of data verification.

QUALITY SERVICE MANAGER

Quality service managers were supposed to monitor the database on a daily basis to ensure accuracy and resolve problems. Additionally, the separate Customer Service and Inbound databases were supposed to be merged by quality service managers. Instead of continued monitoring and improvement, quality service managers complicated the measurement system by adding unnecessary codes and restrictions. If the different Inbound and CSR databases were combined and the coding system simplified, the result may have been a decreased need for the quality service manager position to exist. However, the continued growth of the measurement system made the position important, thereby reinforcing behaviors that may have led to increased complexity and confusion.

VICE PRESIDENT

Ultimately, the vice president for operations used the reports generated by the measurement system to make business decisions. However, the flawed system led the VP to conclude LIH errors due were not a problem. Because he designed the measurement and incentive systems, the VP received praise for seemingly addressing the costly problem of LIHs. With errors averaging approximately $5 per incident, the resolution of one of the most prevalent errors was a major accomplishment. This resulted in positive consequences such as increased responsibility, career growth, and recognition. Repairing the measurement system after it was obvious that it was flawed would have drawn scrutiny of the unchecked costly errors and social disapproval from superiors.

Suggestions for a New Measurement System

The measurement system was designed to capture errors in the delivery process as they occurred and summarize trends to aid in the resolution of these problems. The other goal of the measurement system was to provide the quantitative basis for the company's incentive system. However, the lack of reinforcement for oversight resulted in invalid data remaining hidden for over 5 years and contributed to bad organizational decisions. Additionally, the contingencies surrounding the measurement system reinforced agents for manipulating the processes for their own benefit and/or the benefits of others. The likelihood that an agent will act in the interest of the organization increases when the organization has information that can be used to verify the behavior of the agents (Eisenhardt, 1989). Thus, the measurement system should be repaired by refocusing on its designed purpose.

SYSTEM RESTRUCTURING

A measurement system needs to capture enough detail to inform business decisions and provide information about employee performance. A redesign

must pinpoint areas where the organization can make improvements. If these specific variables are captured by the system, management of performance improvement becomes more effective. Eisenhardt (1989) explained that when managers and employees are rewarded for outcomes dependent on the same behavior, the conflicts between the two parties are reduced and employees are more likely to act in the interest of the organization. A new single measurement system should be developed based on the mission and goals of the organization first, then the goals of the departments, and individuals should be considered next. The resulting goals (e.g., customer service) should then be used to develop the variables to be assessed.

Ultimately, the variables measured must be under the control of the organization and its agents (Bassi & McMurrer, 2005). The agents who use the system daily (e.g., CSRs and Inbound managers) are experts who should be consulted concerning what aspects of the system work, which codes are actually used, and which could be discarded. Consequently, using the agents in system redesign will increase buy-in, which may ensure the system will be supported and used properly. Furthermore, this redesign using the agents who will end up using the system will allow more reliability and validity in the application of codes.

Lastly, the system should be flexible and allow for change when needed. It is essential that the organization routinely review the accuracy of the system and the variables captured by auditing the system. The system must be monitored to ensure that it examines the most important variables and can accommodate any new variables needed. However, these changes should not be at the discretion of any one user. Thus, access to system design features should be limited. By creating an avenue for evaluation and improvement, the continued relevance of the measurement system can be maintained in the face of changes (Garnego et al., 2005).

PROCESS-BASED CHANGE

In the failed measurement system, driver errors could be recorded in 2 systems with almost 250 choices of codes to apply between them. The codes were housed in separate systems and applied by separate agents and had no tangible link to one another. Codes were often redundant within a single system and even more so between the systems. Codes were then merged to form a single report used to make business decisions. This report was mainly informed by the Customer Service database and was supplemented by Inbound data when errors were not captured wholly by the Customer Service data.
In a new system, all agents would enter error codes into the same system. The agents would choose from a number of error code options limited to only the most relevant. Multiple codes should be captured in one problem slip. Additional important information (i.e., agent responsible for error, furniture brand or type) could be made selectable in a searchable format. This would allow for the information needed to be more easily identified through searches without complicating the error coding process.

TRAINING

Individuals who use the system and apply the codes need to be aware of how it is to be used and what it was supposed to accomplish. Technical training will provide information required to use the system (Lewis et al., 2004), such as standardized definitions and examples of each code. Training will increase fluency in application and increase reliability and validity.

AUDITING/ACCOUNTABILITY

Coding inaccuracies were overlooked in the audit process of the current measurement system. According to Power (2004), auditing is the organizational equivalent of scientific replication in that it serves as "institutional revisiting of performance measures" (p. 770). The auditing system needs to be reworked to reduce the aversive costs associated with the process. The existence of two databases measuring the same errors resulted in unnecessary redundancies and response cost in the auditing process. Any measurement system should compile one database and build information technology to automate the audit process. The management team should reinforce frequent auditing, perhaps making it a weekly instead of a quarterly process. The reduced lag time will reduce the number of records to be audited at a given time. Moreover, the company could designate auditing as the primary responsibility of a clerical employee.

FEEDBACK

With enhanced auditing, frequent feedback could be provided. Also, a report on the accuracy of code entry would allow the company to reinforce CSRs' coding. A new measurement system should enhance the timeliness and clarity of feedback to drivers as well. Frequently verified error codes could be translated into driver feedback more often than once a quarter. Delivery performance can be made available electronically through cellular phones or computers in the delivery trucks. In addition, weekly reports should be issued, highlighting areas of success and areas in need of improvement. The feedback loop, then, provides an avenue for evaluation and improvement of the system and performance.

CONCLUSION

Purchase decisions are made on expectations rather than complete information, thus the higher a customer's expectation of product quality, the higher the price the customer is willing to pay (Goering, 1985). Indeed, an organization's competitive advantage depends on the quality of its goods and services (Gronroos, 1988). When making purchasing decisions, the two most important factors according to customers are product quality and the company's handling of service failures (Conlon & Murray, 1996). When customers are not satisfied with the service received, they may relay this information to 10–20

people (Mattila, 2001).

For the business reviewed in this case, a defective furniture piece left in a customer's home could influence future purchasing decisions by these high-end customers. Thus, special measurement systems were designed to influence deliverer behavior. However, very real and significant side effects of the faulty measurement system had gone unnoticed.

The most obvious side effect of measurement system failure was the incentive system failure. The manipulation of the system made these incentives ineffective. Drivers were being reinforced; however, it was not contingent on any accurate measure of performance. As Latham and Dossett (1978) explained, reinforcement (i.e., monetary incentives) needs to be contingent on performance to be effective. Instead, the incentive was resulting in increased expenses for the company without the benefit of managing key quality behaviors visible to the customer. The only real consequence of this system was the illusion of improvement.

REFERENCES

Abernathy, W. B. (2000). *Managing without supervision: Creating an organization-wide performance system*. Atlanta, GA: Performance Management Publications.

Abernathy, W. B., Duffy, E. M., & O'Brien, R. M. (1982). Multi-branch, multi-system programs in banking: An organization-wide intervention. In R. M. O'Brien, A. M. Dickinson, & M. P. Rosow (Eds.), *Industrial behavior modification: A management handbook* (pp. 370–382). New York: Pergamon.

Alvero, A. M., Bucklin, B. R., & Austin, J. (2001). An objective review of the effectiveness and essential characteristics of performance feedback in organizational settings (1985–1998). *Journal of Organizational Behavior Management*, *21*, 3–29.

Anderson, B., & Fagerhaug, T. (2002). *Performance measurement explained: Designing and implementing your state-of-the-art system*. Milwaukee, WI: American Society for Quality Press.

Balcazar, F. E., Hopkins, B., & Suarez, Y. (1985/86). A critical objective review of performance feedback. *Journal of Organizational Behavior Management*, *7*, 65–89.

Banker, R. D., Lee, S.-Y., Potter, G., & Srinivasan, D. (1996). Contextual analysis of performance impacts of outcome-based incentive compensation. *Academy of Management Journal*, *39*, 920–940.

Bassi, L., & McMurrer, D. (2005). Learning from practice: Developing measurement systems for managing in the knowledge era. *Organizational Dynamics*, *34*, 185–196.

Bateman, M. J., & Ludwig, T. D. (2003). Managing distribution quality through an adapted incentive program with tiered goals and feedback. *Journal of Organizational Behavior Management*, *23*, 33–55.

Bernstein, D. A., Penner, L. A., Clarke-Stewart, A., & Roy, E. J. (2003). *Psychology*. Boston: Houghton Mifflin.

Brethower, D. M. (1982). The total performance system. In R. M. O'Brien, A. M. Dickinson, & M. P. Rosow (Eds.), *Industrial behavior modification: A management handbook* (pp. 350–369). New York: Pergamon Press.

Brethower, D. M. (2000). A systematic view of enterprise: Adding value to performance. *Journal of Organizational Behavior Management*, *20*(3/4), 165–190.

Brethower, D. M. (2001). Managing a person as a system. In L. J. Hayes, J. Austin, R. Houmanfar, & M. C. Clayton (Eds.), *Organizational change* (pp. 89–105). Reno, NV: Context Press.

Brethower, D. M. (2002). Behavioral systems analysis: Fundamental concepts and cutting edge application. Retrieved February 28, 2003, from http://www.behavior. org/performancemgmt.

Bucklin, B. R., & Dickinson, A. M. (2001). Individual monetary incentives: A review of different types of arrangements between performance and pay. *Journal of Organizational Behavior Management*, *21*, 45–137.

Conlon, D. E., & Murray, N. M. (1996). Customer perceptions of corporate responses to product complaints: The role of explanations. *Academy of Management Journal*, *39*, 1040–1056.

Daniels, A., & Daniels, J. (2004). *Performance management: Changing behavior that drives organizational effectiveness*. Atlanta, GA: Performance Management Publications.

Dickinson, A. M., & Gillette, K. L. (1993). A comparison of the effects of two individual monetary incentive systems on productivity: Piece rate pay versus base pay plus monetary incentives. *Journal of Organizational Behavior Management*, *14*, 2–82.

Eisenhardt, K. M. (1985). Control: Organizational and economic approaches. *Management Science*, *31*, 134–149.

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, *14*, 57–74.

Flamholtz, E. G. (1979). Effective organizational control: A framework, applications and implications. *European Management Journal*, *14*, 596–611.

Gaetani, J. J., Hoxeng, D. D., & Austin, J. T. (1985). Engineering compensation systems: Effects of commissioned versus wage payment. *Journal of Organizational Behavior Management*, *7*, 51–63.

Garnego, P., Biazzo, S., & Bitici, U. S. (2005). Performance measurement systems in SMEs: A review for a research agenda. *International Journal of Management Reviews*, *7*, 25–47.

Glenn, S. S. (1988). Contingencies and metacontingencies: Toward a synthesis of behavior analysis and cultural materialism. *The Behavior Analyst*, *11*(2), 161–179.

Goering, P. A. (1985). Effects of product trial on customer expectations, demand, and prices. *Journal of Consumer Research*, *12*, 74–82.

Goomas, D. T., & Ludwig, T. D. (2007). Enhancing incentive programs through proximal goals and immediate feedback: Engineered labor standards and technology enhancements in stocker replenishment. *Journal of Organizational Behavior Management*, *27*, 33–68.

Gormezano, I., & Wasserman, E. A. (1992). *Learning and memory*. Hillsdale, NJ: Erlbaum.

Gronroos, C. (1988). Service quality: The six criteria of good perceived service quality. *Review of Business*, *9*, 10–13.

Hall, F. S. (1975). Organizational goals: The status of theory and research. In J. L. Livingstone (Ed.), *Managerial accounting: The behavioral foundations* (pp. 1–32). Columbus, OH: Grid.

Harrington, H. J., & McNellis, T. (2006). *Mobilizing the right lean metrics for success*. Retrieved September 2, 2007, from http://www.qualitydigest.com/may06/articles/02_article.shtml

Honeywell-Johnson, J. A., McGee, H. M., Culig, K. M., & Dickinson, A. M. (2002). Different effects of individual and small group incentives on high performance. *The Behavior Analyst Today*, *3*(1), 88–103.

Jensen, M., & Meckling, W. (1976). Theory of the firm: Managerial behavior, agency costs, and ownership structure. *Journal of Financial Economics*, *3*, 305–360.

Katz, D., & Kahn, R. L. (1966). *The social psychology of organizations*. New York: Wiley.

Kula, W. (1986). *Measures and men*. Princeton, NJ: Princeton University Press.

Lanyon, R. I., & Goodstein, L. D. (1997). *Personality assessment*. Oxford, England: John Wiley & Sons.

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.

Lewis, P. S., Goodman, S. H., & Fandt, P. M. (2004). *Management: Challenges for tomorrow's leaders*. Mason, OH: South-Western.

Ludwig, T. D., & Goomas, D. T. (2007). Performance, accuracy, data delivery, and feedback methods in order selection: A comparison of voice, handheld, and paper technologies. *Journal of Organizational Behavior Management*, *27*, 69–107.

Malott, M. E. (2003). *Paradox of organizational change*. Reno, NV: Context Press.

Malott, R. W. (1974). A behavioral systems approach to the design of human services. In D. Harshbarger & R. F. Maley (Eds.), *Behavior analysis and systems analysis: An integrative approach to mental health programs*. Kalamazoo, MI: Behaviordelia.

Malott, R. W. (1999). *Behavior analysis as world view: A worktext*. Retrieved January 8, 2008, from http://dick-malott.com

Mattila, A. S. (2001). The impact of relationship type on customer loyalty in a context of service failures. *Journal of Service Research*, *4*, 91–101.

Mitnick, B. M. (1992). The theory of agency and organizational analysis. In N. E. Bowie & R. E. Freeman (Eds.), *Ethics and agency theory* (pp. 75–96). New York: Oxford University Press.

Mitnick, B. M. (1998). Agency theory. In R. E. Freeman & P. H. Werhane (Eds.), *The Blackwell encyclopedic dictionary of business ethics* (pp. 12–15). Malden, MA: Blackwell.

Moe, T. M. (1984). The new economics of organization. *American Journal of Political Science*, *28*, 739–777.

Power, M. (2004). Counting, control and calculation: Reflections on measuring and management. *Human Relations*, *57*(6), 765–782.

Pritchard, R. D., Jones, S. D., Roth, P. L., Stuebing, K. K., & Ekeberg, S. E. (1988). Effects of group feedback, goal setting, and incentives on organizational productivity. *Journal of Applied Psychology*, *73*(2), 337–358.

Pritchard, R. D., Jones, S. D., Roth, P. L., Stuebing, K. K., & Ekeberg, S. E. (1989). The evaluation of an integrated approach to measuring organizational productivity. *Personnel Psychology*, *42*, 69–115.

Rogers, C., & Skinner, B. F. (1956). Some issues concerning the control of human behavior. *Science*, *124*, 1057–1066.

Rummler, G. A. (2004). *Serious performance consulting according to Rummler*. Silver Spring, MD: International Society for Performance Improvement.

Shapiro, S. P. (2005). Agency theory. *Annual Review of Sociology*, *31*, 263–284.

Sieber, S. (1981). *Fatal remedies: The ironies of social intervention*. New York: Plenum Press.

Skinner, B. F. (1953). *Science and human behavior*. New York: Free Press.

Sulzer-Azaroff, B. (2000). Of eagles and worms: Changing behavior in a complex

world. *Journal of Organizational Behavior Management*, *20*(3/4), 139–163.

Tangen, S. (2004) Performance measurement: From philosophy to practice. *International Journal of Productivity and Performance Management*, *53*, 726–737.

Welsch, D., Luthans, F., & Sommer, S. (1993). Managing Russian factory workers: The impact of U.S.-based behavioral and participative techniques. *Academy of Management Journal*, *36*, 58–79.

White, H. (1985). Agency as control. In J. Pratt & R. Zeckhauser (Eds.), *Principals and agents: The structure of business* (pp. 187–214). Boston: Harvard Business School Press.

Williamson, O. E. (1975). *Markets and hierarchies: Analysis and antitrust implications.* New York/London: Free Press.