PRACTITIONER'S CORNER

Measuring Customer Service: Georgia's Local Government Mystery Shopper Program

Mark D. Bradbury and Richard L. Milford

FOR LOCAL GOVERNMENTS, performance measurement efforts typically focus on workload or output measures. One aspect of organizational performance that receives less attention is the actual process of service delivery. The quality of customer service provided by frontline staff to citizens should be important to administrators because "courtesy and friendliness become important not as ends in themselves, but because customers partially conflate delight at courteous and friendly treatment with the actual quality of service" (Fountain 2001, 58).

Poisant (2002) suggests that to analyze customer, or citizen, experiences, an organization must put itself in citizens' shoes and ask questions about how they are greeted, the appearance of frontline staff, and the aesthetics of the office place (Poisant 2002, 106). Mystery shopping is a participant-observer evaluation method that can be used to measure the quality of customer service provided by frontline staff as it unfolds (Grove and Fisk 1992). Mystery shopping uses a trained observer posing as a client or customer to contact a government agency seeking service or information. The technique focuses on "activities and procedures that do or do not occur rather than gathering opinions about the service experience" (Wilson 2001, 725).

Thus, mystery shopping provides an alternative to after-the-fact customer surveys or other feedback techniques because it focuses on the process of service delivery rather than on the outcome of a service encounter.

The validity of the mystery shopper technique hinges on the employee believing that the observer is in fact a real customer, not an "undercover" evaluator. To both protect the anonymity of the shopper and to increase the believability of the interaction, mystery shoppers follow a script, or scenario, to solicit responses from the employee. Immediately after the interaction, the evaluator rates the quality of customer service received based on an established set of criteria. Following Patton's (1990) framework of observational methods, mystery shoppers can be thought of as participant evaluators who make multiple, covert observations that focus on the service-delivery aspect of programmatic effectiveness.

Mystery shopping is best used strictly for research on general patterns and quality of service delivery rather than for assessing the performance of individual employees. The European Society for Opinion and Marketing Research (ESOMAR) cautions that "researchers must not undertake Mystery Shopping studies which would be used for

nonresearch purposes, such as checking the performance of specific identified individuals for possible disciplinary action" (ESOMAR 2002). Additionally, individual staff members should not be identified by name in any report, and data should only be reported in the aggregate.

This article focuses on an innovative program designed to systematically measure customer service for local governments in Georgia. Coordinated by the Carl Vinson Institute of Government at the University of Georgia, the Local Government Mystery Shopper Program offered local governments a systematic, third-party evaluation of customer service. The program attempted to replicate typical citizen–service administrator encounters by analyzing both walk-in and telephone interactions and by using scenarios that were specific to the functions of local government departments. As a result, the Mystery Shopper Program provided ongoing customer service evaluation data that were used by participating local governments to improve organizational performance and to tailor customer service training for their employees.

Before discussing the specifics of the Mystery Shopper Program, it is important to provide the context for the types of citizen–service administrator interactions that the program replicated. Based on Goodsell's discussion (1981, 4–7), the core elements of the interaction are as follows. The interaction was between two people, the service administrator and the mystery shopper, and was purposive rather than random or by accident. The interaction replicated one between the public and administrative personnel rather than between two administrators or with elected officials. The encounter was initiated by a citizen rather than by the government, as in a code violation or arrest. All of the encounters occurred at government offices rather than at citizens' residences or in a public place. Finally, scenarios were designed to create interactions that resulted in an exchange of information and were routine, brief, and one-shot in nature. Overall, these characteristics were meant to approximate typical citizen–service administrator interactions.

The Mystery Shopper Program

The desire for a customer service measurement program came from discussions among faculty at the Carl Vinson Institute of Government and city and county managers throughout Georgia. These local government leaders wanted to know how average citizens are treated when they come into contact with local government service administrators. A mystery shopping program was adopted because it can provide up-to-date information detailing what actually occurs at the key moment of customer interaction. A steering committee consisting of local government officials, Vinson Institute faculty and staff, and customer service experts was established to design key components of the program.

Georgia's Local Government Mystery Shopper Program operated from the summer of 1999 through December 2001. City and county governments signed up for the program every six months. Governments enrolled anywhere from 1 to more than 20 departments. Each participating department was "shopped" 10 times during the six-month shopping period. Each six-month period was broken down into two three-month subperiods, with five "shops" per department occurring in each subperiod. The purpose of these subperiods was to spread out the shops to the greatest possible extent over the course of the six-month period. This procedure helped prevent all 10 shops from occurring during a particularly hectic or slow period for the department being shopped and thus provided a more balanced evaluation of typical customer service experiences. Overall, 30 cities and 12 counties of various sizes participated in the program, and a total of 4,018 shopper evaluations were completed.

Trained Evaluators

The Vinson Institute contracted with Shop'n Chek, Inc., of Atlanta in the fall of 1998 to

conduct the mystery shopping. Shop'n Chek has provided mystery shopping services for restaurants, retail companies, manufacturing, utilities, and, to a lesser extent, government since 1972. The company's network of trained, objective evaluators provided the necessary pool of shoppers to serve local governments throughout the state.

Shop'n Chek's initial response was one of reluctance. The company had worked with a few individual local governments in the past and found those relationships to be unprofitable because of the relatively small number of shops per client. This concern was alleviated by having the Vinson Institute act as the "home office," or middleman. Interested local governments signed up with the Vinson Institute to have certain departments shopped, and these shops were collectively treated as one account with Shop'n Chek.

Shop'n Chek was also responsible for training shoppers to be objective and consistent evaluators. In addition to this fundamental training, the nature of mystery shopping requires shoppers who are "objective, smart, and able to think on their feet" (Leeds 1995). For the Mystery Shopper Program, both the contact people at Shop'n Chek and the actual shoppers had to be educated about the differences between county and city government responsibilities as well as the functions of individual departments therein. The need to educate Shop'n Chek and its shoppers regarding local government functions underscores an important point about public-sector service delivery. Citizens who contact government with a question or problem often do not know how to communicate their needs in a clear, concise manner or know which department they need to contact. This problem highlights the need for courteous and patient customer service on the part of governmental personnel.

Appropriate Locations

To evaluate the quality of customer service, an organization must have a requisite level of contact with outside vendors or clients.

For the Mystery Shopper Program, a local government office had to have a high level of citizen contact, not merely serve internal customers (e.g., an information technology or purchasing department). Moreover, the steering committee addressed the issue of whether participating governments should inform their employees that mystery shopping was going to be used to evaluate departmental performance. The standard approach is for managers to inform employees before any mystery shopping occurs (see Leeds 1995; Wilson 2001; ESOMAR 2002). The Georgia Program was designed so that the results could not be used to evaluate the performance of individual personnel. Thus, the degree of intrusion was reduced, as was the risk that the results of the mystery shopping could harm an individual employee. Consequently, a middle-ground option was forged. The decision was left to participating governments whether or not to inform their frontline staff that mystery shopping was to occur. Although no records were kept, communications with participating governments indicated that many chose to inform their employees prior to the commencement of mystery shopping.

Functional Scenarios

In addition to a trained, objective evaluator, a mystery shopping evaluation requires a credible scenario that replicates a common customer-employee interaction. A high-quality scenario has several characteristics. First, the scenario must approximate a common interaction between a citizen and a government employee, not only for validity purposes but also to protect the anonymity of the shopper. Second, the scenario must require more than a dichotomous "yes" or "no" answer. That is, the government employee must be prompted to provide some sort of information to the shopper. Additionally, the scenario must present a question that can be answered reasonably by the employee whom the shopper encounters, not result in an automatic referral to another department or office. Most important,

the scenario must engage the employee in some interaction with the shopper.

The steering committee identified 19 functions commonly performed by local governments (see Table 1). Over the course of the program, nearly 70 percent of the departments that were shopped could be categorized into one of six functional areas: Recreation/Community Center (16 percent), Utilities (14 percent), Building Inspection/Zoning (12 percent), Public Works/Solid Waste (9 percent), Police/Fire (9 percent), and Human Resources (8 percent). At least two scenarios were developed for each functional category (see Table 2 for examples). For walk-in evaluations, it was possible for the same shopper to make more than one visit to a department in the six-month period. Therefore, shoppers were instructed to select a scenario from the lists to vary the evaluation experiences and decrease the possibility of the shopper's anonymity being compromised.

Customer Service Standards

The literature on mystery shopping assumes that organizations have customer service standards in place. Notably, Wilson (2001, 723)

Table 1. Functional Service Categories

	Percent
A: Building Inspection/Zoning and Planning Departments	12
B: Animal Control or Animal Shelters	2
C: Utilities Departments (water, cable, electricity, gas, or sewer)	14
D: Administration	6
E: Finance	3
F: Purchasing	2
G: Human Resources	8
H: Public Works/Roads and/or Solid Waste	9
I: Tax Assessor or Tax Commissioner	6
J: Police/Fire Departments	9
K: Courts	5
L: Recreation/Community Center Community Services/Community Service Extension	16
M: Voter Registration	1
N: Libraries/Conventions Center Services	2
O: Downtown Development	3
P: Golf	2
Q: Parking/Transit	2
R: Government Information Systems	<1
S: Emergency Management	<1

Table 2. Examples of Function-Specific Scenarios

Category A: **Building Inspection/Zoning and Planning Departments**

What permits do I need for residential construction, and how do I apply for them?

or

When is the next planning and zoning meeting and is there an agenda available?

Category C: **Utilities Departments**

What rates do you charge for (water and sewer, cable television, electricity, gas), and how long does it take for service to be turned on?

or

If you are an actual customer in (city/county), you can ask any question that you may have about your bill.

Category G: **Human Resources**

How do I find out what jobs are available?

or

How do I apply for a job you have listed?

Category H: **Public Works/Roads and/or Solid Waste**

How do I dispose of (tires, oil, paint), and where is the location?

or

What streets are scheduled to be paved this year?

Category J: **Police/Fire Departments**

I work with a (examples: senior citizens group, youth church group). What kind of public educational programs do you have?

or

What kinds of residential fire inspection services do you offer?

Category L: **Recreation/Community Center/ Community Services/ Community Service Extension**

What athletic leagues do you have?

or

Where are your parks located, and what are your hours?

posits that "head office personnel and senior management invariably set service standards." This situation was not the case, however, for many local governments in Georgia. The vast majority of local government employees who were shopped had not received previous customer service training. Thus, the program provided a picture of what customer service behaviors could be expected from employees in the absence of formal training or direction from superiors.

Similarly, there is no agreed-upon list of behaviors that constitute "good" customer service. Consequently, the steering committee developed a standard instrument for measuring customer service for local governments. Most of the questions were adapted from customer service experiences in the private sector. The process of selecting the behaviors to be observed and graded was inherently subjective but based on the varied experiences of the steering committee members. Although participating governments periodically requested that certain less important measures be eliminated from use when their departments were shopped, the agreement between the Vinson Institute and Shop'n Chek called for the use of one uniform evaluation instrument for each type of shop, identified as a walk-in shop and a telephone shop.

Evaluations were available for both walk-in and telephone interactions. Walk-in shoppers collected data on a series of 15 standard customer service measures, including the ease of finding the appropriate office, whether their questions were answered fully, if they were referred to the correct department, and whether the employee was courteous and/or genuinely helpful (Table 3). Telephone shops included many of the same measures, plus data on how many times the telephone rang before being answered and how long the shopper was put on hold (Table 4). Although the measures may seem simplistic, they were designed to represent the behaviors from which impressions are made about the quality of local government service.

The maximum score for each mystery shop was 100, with some aspects of the interaction given greater weight than others (see Tables 3 and 4). The decision of the weights for each measure was made by the steering committee. The weighting of measures, albeit inherently subjective, reflects the committee's perceived importance of a given behavior to overall customer service. Behaviors were dichotomously scored, so the trained shopper assigned, for example, 10 points if the employee was helpful or no points if the employee was not helpful. At the end of each six-month shopping pe-

Table 3: Walk-In Customer Service Criteria and Observations

Customer Service Criteria	Result (%)	Weight
Locating the Department/Office:		
Is the building easily located by address provided?	93	5
Is the office well identified?	90	10
Once inside the Office/Department:		
How long did you have to wait to have someone offer to help you?	43.3 sec.	—
Did the employee:		
Make eye contact?	98	5
Smile and greet you pleasantly?	87	5
Appear neat and well groomed?	98	5
Use your name at any point in the transaction?	36	5
Did the employee:		
Ask questions to clarify your desired needs?	74	10
Answer your questions fully?	85	10
Seem genuinely concerned about providing an answer to your inquiry?	89	10
Refer you to another source?	73	—
Offer to help you with anything else?	52	10
Thank you or give you a pleasant closing statement?	76	5
Overall, was the employee:		
Pleasant and friendly during your entire visit?	92	10
Helpful?	93	10

Table 4: Telephone Customer Service
Criteria and Observations

Customer Service Criteria	Result (%)	Weight
Telephoning the Department/Office:		
How many times did the telephone ring before someone answered?	1.86 rings	—
Were you placed on hold?	25	—
If placed on hold, how long?	40.0 sec.	—
Answering the telephone:		
Did the employee:		
Identify the department?	91	10
Identify himself/herself?	60	10
Offer to help you?	80	5
Use your name at any point in the transaction?	18	5
Did the employee:		
Ask questions to clarify your needs?	62	10
Attempt to answer your questions fully?	78	10
Seem genuinely concerned about providing an answer to your inquiry?	87	15
Refer you to another source?	70	—
Offer to help you with anything else?	39	10
Thank you or give you a pleasant closing statement?	71	5
Overall, was the employee:		
Pleasant and courteous?	91	10
Helpful?	90	10

riod, each participating government received copies of the shoppers' evaluations and a set of tabulated results comparing that government's departments with other participants.

Data

Scores on the 15 elements of quality customer service show the behaviors that employees exhibit in the absence of formal customer service training. Because of the behavioral differences between walk-in and telephone interactions, the evaluation instruments are slightly different for the two types of shops. The results from walk-in shops are therefore analyzed separately from the telephone data.

From July 1999 through December 2001, 1,649 walk-in shops and 2,369 telephone shops were completed. The average score was 84 percent for walk-in shops and 75 percent for telephone shops, suggesting that in the absence of training specifically designed to improve customer service, local government employees tend to provide better customer service to citizens in face-to-face interactions than over the telephone.

Analysis of the scores for behaviors observed for walk-in shops reveals several tendencies (Table 3). Generally, local government buildings were easily located, and offices were well identified. Overall, employees were helpful, pleasant, and friendly. Employees appeared to be neat and well groomed, made eye contact, and to a slightly lesser extent smiled during the transactions. Employees were somewhat less likely to ask questions to clarify the shopper's needs, to appear genuinely concerned with providing an answer to the inquiry, or to fully answer the question. Employees were even less likely to thank the shopper or offer to help with anything else. The lowest scores relate to whether the employee used the shopper's name at any point during the interaction (shoppers were trained to introduce themselves by name at the start of the interaction).

Similar variations were found in the scores from telephone shops (Table 4). The phone rang an average of 1.86 times before it was answered, and 25 percent of calls were placed on hold. Overall, local government employees were helpful, pleasant, and friendly. Although employees consistently identified the department and seemed genuinely concerned with the inquiry, they were much less likely to identify themselves or to use the caller's name at any point during the transaction. Employees were somewhat unlikely to offer to help the caller, to ask questions to clarify the caller's needs, or to fully answer the question. Even fewer employees thanked callers or offered to help with anything else.

A number of behaviors were observed for both walk-in and telephone shops. A com-

parison of the scores on the common measures confirms the observed differences in the behaviors for walk-in shops and telephone shops. The following behaviors were notably more common for walk-in shops compared with telephone shops: stating the customer's name at any point during the transaction (36 percent compared with 18 percent), asking questions to clarify the customer's needs (74 percent compared with 62 percent), and fully answering the question at hand (85 percent compared with 78 percent). Furthermore, employees were more likely to say thank you (76 percent compared with 71 percent) and to offer to help with anything else (52 percent compared with 39 percent) to a customer who is standing before them than to one who is on the phone.

These findings indicate that local government employees tend to provide higher quality customer service in face-to-face interactions with citizens than they do over the telephone. Second, scores were high for attributes related to appearance and pleasantness but comparatively low for behaviors such as thanking the citizen and offering to help with any other issues. These results suggest that social norms related to appropriate dress and demeanor in the workplace are generally adhered to, but training may be necessary to instill specific behaviors associated with excellent customer service.

How Some Governments Are Using the Data to Improve Quality in Customer Service

Some local governments have attempted to "raise the bar" in the provision of customer service. Two participating local governments held retreats with their department heads to analyze the results of the mystery shops and to determine ways to address the problems associated with their customer service. One created a new position of Director of Customer Service to continuously follow up on governmentwide efforts to improve customer service. The Vinson Institute provided this

government with customized customer service training based on the 15 customer service measures. The second government established standards of customer service based on the program's measures. The department heads communicated the newly established standards to their staff and established expectations for employees on how their citizens and customers should be treated.

A third participating government showed dedication to improved customer service by using multiple programs. As part of a comprehensive strategic planning effort, it conducted citizen surveys to measure customer satisfaction with the government's four utilities and general services. Baselines were established to measure customer service improvements. The government then participated in the Mystery Shopper Program to collect ongoing data to guide its training and improvement efforts.

Further Improvements to the Program

The Georgia Local Government Mystery Shopper Program was a unique and challenging endeavor, and the following modifications should be considered if the program is replicated. First, the training regimen for mystery shoppers must be continuously updated and revised as new technologies are integrated into the public-sector workplace. Shoppers need to be trained to think on their feet and deal with unexpected circumstances.

Second, the scoring system used for both walk-in and telephone shops in the Mystery Shopper Program was designed to award all or no points for a given characteristic. Given the inherent subjectivity of such a measure, this point system could be modified to allow for a range or scale of points to be awarded. On a scale for politeness, for example, a shopper could award 10 points for excellent, 8 points for good, 5 points for acceptable, 2 points for poor, and 0 points for unacceptable (Wilson 2001, 726–27).

Third, the contractual arrangement between the Vinson Institute and Shop'n Chek

required the use of a uniform evaluation instrument for all walk-in and telephone shops. Over time, participating governments expressed interest in modifying the instrument to emphasize behaviors that were of greater importance and remove those deemed less important to quality customer service. The problem was that different governments wanted to modify the instrument in different ways. Ideally, such modifications would be possible, resulting in an evaluation instrument that is uniquely tailored to the customer service priorities of each participating government.[1] Finally, participating governments requested that the scenarios be modified to specifically address the services delivered by their departments. Similar to the problem of having uniquely tailored evaluation instruments, the contractual structure of the program did not allow for a different slate of scenarios to be used for each participating government.

Although mystery shopping is a valuable tool for assessing the quality of customer service, its influence can be short-lived. Without an ongoing commitment from management to reinforce service standards, "the novelty of being shopped can wear off, leaving personnel complacent about their service and lacking motivation to take steps to improve it further" (Wilson, 2001, 732). The technique of mystery shopping perhaps works best when coupled with other measurement devices, such as general customer service surveys, and followed up with a concentrated training component for management and staff alike.

The importance of quality customer service in the public sector cannot be overstated. Because any given citizen only periodically comes into contact with government personnel, those few experiences are greatly influential on that citizen's perception of the quality of local government services. For this reason, the Mystery Shopper Program serves as a valuable measurement tool for continuously assessing employee-customer interactions. If the program is replicated, participants are encouraged to take responsibility for seeking ways to use the data produced by the shops to routinely address areas needing improvement.

Mark D. Bradbury *is a faculty member in the Masters of Public Administration program at Binghamton University, New York. His research interests include public human resource management, public management, and public-sector ethics. His work has appeared in the* Georgia Law Review *and the* Journal of Health and Human Services Administration.

Richard L. Milford *is a public service associate with the Carl Vinson Institute of Government at the University of Georgia. His work involves program evaluation and technical assistance with state and local governments. Particular studies have focused on workforce training and development efforts of higher education institutions.*

Note

1. For an example of a slate of measures specifically designed for a client, see the discussion of the London Underground public transport network in Wilson and Gutmann (1998).

References

ESOMAR. 2002. *Guidelines on mystery shopping*. http://www.esomar.nl/guidelines/mysteryshopping.htm.

Fountain, Jane E. 2001. Paradoxes of public sector customer service. *Governance* 14, no. 1: 55–73.

Goodsell, Charles T. 1981. *The public encounter*. Bloomington: Indiana University Press.

Grove, S. J., and R. Fisk 1992. Observational data collection methods for services marketing: An overview. *Journal of the Academy of Marketing Science* 20: 217–24.

Leeds, Barry 1995. Mystery shopping: From novelty to necessity. *Bank Marketing* 27, no. 6: 17–22.

Patton, Michael Quinn 1990. *Qualitative evaluation and research methods*. Newbury Park, CA: Sage.

Poisant, Jim 2002. *Creating and sustaining a superior customer service organization*. Westport, CT: Quorum Books.

Wilson, Alan M. 2001. Mystery shopping: Using deception to measure service performance. *Psychology & Marketing* 18, no. 7: 721–34.

Wilson, Alan M., and Justin Gutmann. 1998. Public transport: The role of mystery shopping in investment decisions. *Journal of the Market Research Society* 40, no. 4: 285–93.