



“You can expect what you inspect.”

Introduction

This management philosophy is as true today as it was 50 years ago when W. Edwards Deming coined it. Managers of the customer experience have several tools available to them to inspect or monitor the customer experience. However, when it comes to monitoring employee behaviors - service and sales behaviors that drive customer experience success - no tool is better suited for that objective than mystery shopping.

These behaviors – the ones that matter most - are those which drive either purchase intent or customer loyalty.

Mystery shopping programs, when administered in accordance with certain mystery shopping best practices, not only test for the presence of service behaviors, but identify which sales and service behaviors matter most. These behaviors – the ones that matter most - are those which drive either purchase intent or customer loyalty. Mystery shopping provides a vehicle to not only measure but motivate these key behaviors. Central to the success of any customer experience initiative is understanding and adhering to certain best practices. This white paper advances several mystery shopping best practices.

Central to monitoring the customer experience is an understanding of the brand-customer interface. At the center of the customer experience are the various channels which form the interface between the customer and the brand. Together, these channels define the brand more than any external messaging. Different research tools have different research purposes. Some are designed to monitor the customer experience from the customer side of this interface; others, like mystery shopping monitor it from the brand side of the interface. Best in class mystery shopping programs focus on the behavioral side of the equation, answering the question: are our employees exhibiting appropriate sales and service behaviors and are these behaviors the ones that matter?

Types of Mystery Shopping

Before discussing best practices in mystery shopping, it is instructive to consider how brands use mystery shopping to measure and motivate the desired customer experience. Just about any channel in the brand-customer interface can be shopped at any point in the customer journey.

Some of the types of shops include:



In-Person: While distribution channels shift to more self-administered on-line channels, in many industries the in-person channel continues to be the embodiment of the brand – central to a multichannel strategy. This role will put new pressures on store personnel as brand advocate. In-person mystery shopping evaluates and motivates sales and service behaviors as part of this role.



Contact Center: Contact center mystery shopping provides managers a unique opportunity to evaluate the customer experience using predetermined scenarios. Most contact centers employ call monitoring to evaluate agent performance. Best in class mystery shopping programs augment call monitoring by giving managers a tool to present specific scenarios to agents to test their performance.



Internal Shops: Internal shops evaluate service provided to internal customers to identify internal bottlenecks which may hinder the ability to provide optimal customer service.



Web/Mobile Shops: Across many industries, self-administered channels are increasingly becoming key to opening and deepening the customer relationship. Mystery shopping website and mobile channels provide managers tools to test ease of use, navigation and the overall customer experience of online and mobile channels.



Life Cycle Shops: Life cycle mystery shops are designed to evaluate the customer experience through the entire customer journey across a variety of delivery channels, and a spectrum of transactions, over an extended period of time.



Competitive Shops: Shopping competitors allows customer experience managers to benchmark their brand-customer interface relative to their competitors.

Define Objectives

The first step in building a best in class mystery shop program is defining your objectives. Defining research objectives prior to making any other decisions about the program will ensure your program starts right, stays on track, on budget, and produces positive results. The mystery shop best practice in defining objectives for a program is a fairly simple process. First, generate a list of specific behavioral expectations you have of your employees.

What Do You Expect?

Ask yourself what sales and service behaviors you expect from employees. This list of behaviors is going to vary broadly from industry-to-industry, channel-to-channel, and brand-to-brand. Some of the questions you might ask yourself look like this:

- What specific service behaviors do we expect?
- When greeting a customer, what specific behaviors do we expect from staff?
- When meeting with customers after the greeting, what specific behaviors do we expect?
- If a phone interaction, what specific hold/transfer procedures do we expect (for example asking to be placed on hold, informing customer of the destination of the transfer)?
- Are there specific profiling questions we expect to be asked? – If so, what are they?
- What closing behaviors do you expect? How do you want employees to ask for the business?
- At the conclusion of the interaction, how do you want the employee to conclude the conversation or say goodbye?
- Are there specific follow-up behaviors that you expect, such as getting contact information, suggesting another appointment, or offering to call the customer?
- What other specific behaviors do we expect?

Map Expectations to the Shop Questionnaire

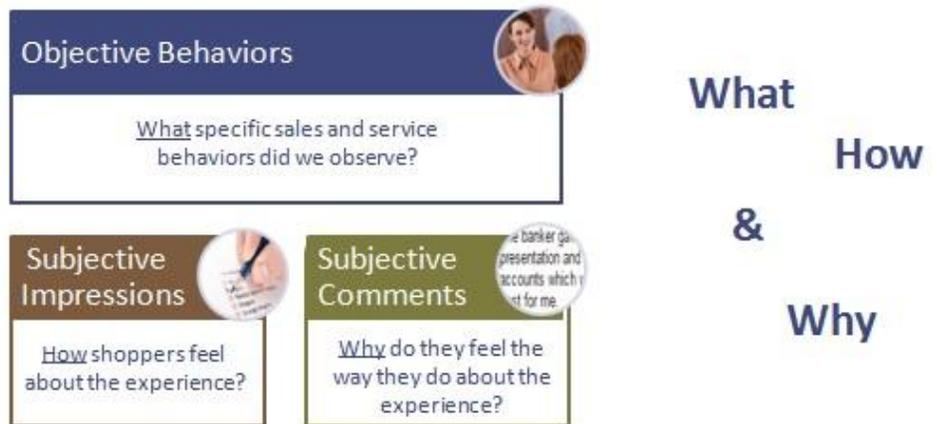
Once you have developed a list of the specific behaviors you expect, the next step is to map each of your behavioral expectations to a question or set of questions on the mystery shop questionnaire. Remember, these behaviors must be specific, objective and observable.

Questionnaire Design Keep it Simple

Often mystery shopping programs are designed by committee which can result in an overly complicated and cumbersome program. Unrealistic scenarios combined with long, overly complex questionnaires result in frustration for mystery shopper, mystery shop provider and the end client. In such cases the likelihood of shopper exposure is increased and the accuracy of the observations suffers. Keep it simple - simpler designs work better and provide more value.

Anticipate the Analysis Finally, identify what specific desired outcome you want from the customer as a result of the experience. Do you want the customer to purchase something? Do you want them return for another purchase? The answer to this question will anticipate the analysis and build in mechanisms for Key Driver Analysis to identify which behaviors are more important in driving this desired outcome – behaviors that matter.

What, How & Why A best practice in mystery shop questionnaire design is to include observations of objective behaviors, subjective impressions and comments. Each of these serves a specific purpose in identifying the service behaviors that matter - behaviors which drive profitability. Together, these three elements of questionnaire design reveal the “what”, “how” and “why” of the customer experience.



What

Objective Behaviors: Observations of objective behaviors form the backbone of best in class mystery shops. These observations identify what specific sales and service behaviors were observed. Mystery shopping is primarily an observational form of research, and as such, a best practice in mystery shopping is to focus on observations of specific objective and observable behaviors. These objective observations serve two purposes: First, they measure and motivate expected sales and service behaviors.

Second, they serve as a foundation for Key Driver Analysis, where the other two subjective elements of the questionnaire are used to determine the relationship between employee behaviors and a desired outcome, such as purchase intent or customer loyalty.

How

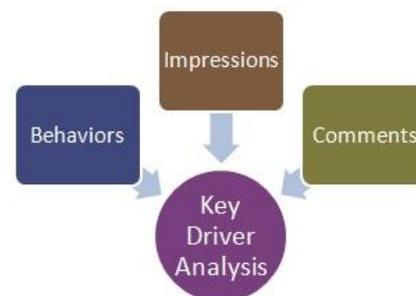
Subjective Impressions: Subjective impressions are primarily captured through scientifically designed and strategically selected rating scales. These questions reveal how the shopper felt about the experience. They add both a quantitative and qualitative perspective to the objective behaviors observed and provide a basis for interpretation of not only individual shops, but also an analytical means to determine the relationship between each service behavior and the desired outcome. *We will explore this in more detail in a discussion of Key Driver Analysis on page 9.*

Why

Subjective Comments: Beyond measuring what behaviors were observed and how the shopper felt about the experience, open-ended comments capture why shoppers felt the way they did about the experience. While objective behaviors are the backbone of the shop, many of Kinēsis' clients consider these comments the heart of the shop, providing a qualitative texture to understand specifically what the shopper felt about the experience. They not only serve as a framework for understanding each shop individually, but provide raw material for content analysis to determine key qualitative key drivers of the desired outcome (purchase intent and customer loyalty).

Anticipate the Analysis

A best practice in mystery shopping program design is to anticipate the analysis. Together, these three design elements provide input into Key Driver Analysis techniques to identify key sales and service drivers of purchase intent and loyalty – behaviors that matter.



Scoring

Most mystery shopping programs score shops according to some scoring methodology to distill the mystery shop results down into a single number.

Scoring methodologies vary, but the most common methodology is to assign points earned for each behavior measured and divide the total points earned by the total points possible, yielding a percent of points earned relative to points possible. It is a best practice in mystery shopping to calculate the score for each business unit independently (employee, store, region, division, corporate).

Not all Behaviors are Equal

Some behaviors are more important than others. As a result, best in class mystery shop programs weight behaviors by assigning more points possible to those deemed more important. Best practices in mystery shop weighting begin by assigning weights according to management standards (behaviors deemed more important, such as certain sales or customer education behaviors), or according to their importance to a desired outcome such as purchase intent or loyalty. Service behaviors with stronger relationships to the desired outcome, identified through Key Driver Analysis, receive stronger weight. *Again, see page 9 for a discussion of Key Driver Analysis.*

Don't Average Averages!

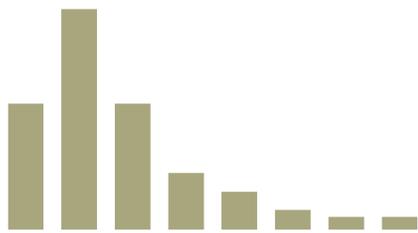
It is a mistake to calculate business unit scores by averaging unit scores together (such as calculating a region's score by averaging the individual stores or even shop scores for the region). This will only yield a mathematically correct score if all shops have exactly the same points possible, and if all business units have exactly the same number of shops. However, if the shop has any skip logic, where some questions are only answered if specific conditions exist, different shops will have different points possible, and it is a mistake to average them together. Averaging them together gives shops with skipped questions disproportionate weight. Rather, points earned should be divided by points possible for each business unit independently. Just remember – don't average averages!

What Is A Good Score?

This is perhaps the most common question asked by mystery shop clients – one for which there is no simple answer. It amazes me how many mystery shop providers I've heard pull a number out of the air, say 90%, and quote that as the benchmark with no thought given to the context of the question. The fact of the matter is much more complex. Context is key. What constitutes a good score varies dramatically from client-to-client, program-to-program based on the specifics of the evaluation. One program may be an easy evaluation, measuring easy behaviors, where a score must be near perfect to be considered “good” – others may be difficult evaluations measuring more difficult behaviors, in this case a good score will be well below perfect. The best practice in determining what constitutes a good mystery shop score is to consider the distribution of your shop scores as a whole, determine the percentile rank of each shop (the proportion of shops that fall below a given score), and set an appropriate cut off point. For example, if management decides the 60th percentile is an appropriate standard (6 out of 10 shops are below it), and a shop score of 86% is in the 60th percentile, then a shop score of 86% is a “good” shop score.

Context is key. What constitutes a good score varies dramatically from client-to-client, program-to-program based on the specifics of the evaluation.

Work toward a Distribution



When all is said and done, the product of a best in class mystery shop scoring methodology will produce a distribution of shop scores, particularly on the low end of the distribution. Mystery shop programs with tight distributions around the average shop score offer little opportunity to identify areas for improvement. All the shops end up being very similar to each other, making it difficult to identify problem areas and improve employee behaviors. Distributions with scores skewed to the low end, make it much easier to identify poor shops and offer opportunities for improvement via employee coaching. If questionnaire design and scoring create scores with tight distributions, consider a redesign.

Sample Plan

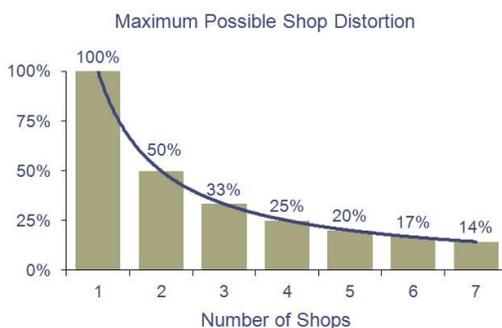
Decisions regarding the number of shops are primarily driven by budgetary resources available and the level of statistical reliability required.

Reliability at Individual or Store Level

The most appropriate measure of reliability at the individual or store level is maximum possible shop distortion (MPSD). Given that shops are snapshots of specific moments in time, it is possible for unique events to influence the outcome of any one shop. It is possible, therefore, that the experience observed by the mystery shopper is not representative of what normally happens. Consider the following examples: a retail location is shopped hours after it was held up, or a bank teller is shopped on the day after her child was up sick all night, or a server at a restaurant just had an extremely bad day. In each of these cases, it is possible these external events impacted employee performance and the customer experience.

How do we know if the experience is typical or not?

Maximum possible shop distortion is the maximum influence any unique event can have on a set of shops to an individual or location.



With one shop to a given location, we do not know if it is typical or not; we only have one data point, so the MPSD is 100%. It is possible the experience is not representative of what is typical. With two shops, the MPSD is 50%. If there are discrepancies within the shops, we do not know which is normal and which is the outlier. With three shops, we now have potentially two shops to point to the outlier (MPSD 33%). The MPSD continues to decline with each additional shop.

As this graph illustrates, maximum possible shop distortion begins to flatten out relative to the incremental program cost as we approach 3 to 4 shops per store. This is where ROI in terms of improved reliability is maximized.

Program Launch & Fielding

Obtain Buy-In From the Front-Line

When mystery shopping initiatives fail to meet their potential, it is often because the people who are accountable for the results — front-line employees, supervisors, store managers, and regional managers — were never properly introduced to the program. As a result, there may be internal resistance, creating an unnecessary distraction from the achievement of the company's service improvement goals. A mystery shopping best practice is to ensure employees throughout the organization are fully informed and have bought into the mystery shopping program before it is launched. Pre-launch efforts should include: the specific behaviors expected of customer facing employees, a copy of the mystery shop questionnaire, training on how to read mystery shopping reports, how to use the information effectively, and how to set goals for improvement.

Provide Adequate Internal Administration

A best practice in mystery shop program design is to anticipate the amount of administration necessary to run a successful mystery shopping program. It requires a strong administrator to keep the company focused and engaged, and to make sure that recalcitrant field managers are not able to undermine the program before it stabilizes and begins to realize its potential value.

Provide a Fair & Firm Dispute Process

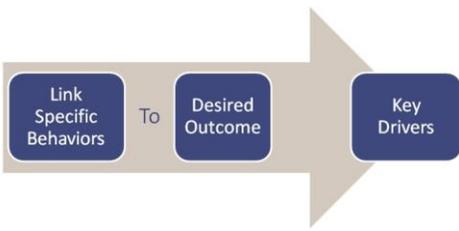
Disputed shops are part of the process. Mystery shops are just a snap shot in time, measuring complex service interactions. As a result, there may be extenuating circumstances that need to be addressed, or questions about the quality of the mystery shopper's performance that require both a fair and firm process to dispute shop scores. Fairness is critical to employee buy-in and morale. Firmness is required to keep the number of shop disputes in check, and cut down on frivolous score disputes.

The specifics of the dispute process will depend on each brand's culture and values. Here are some ways a fair and firm best in class mystery shop dispute process can be designed:

Arbitration: Most brands have a program manager or group of program managers acting as an arbitrator of disputes and ordering reshops or adjusting points to an individual shop as they see fit. The arbiter of disputes must be both fair and firm, otherwise, employees and other managers will quickly start gaming the system, bogging the process down with frivolous disputes.

Fixed Number of Challenges: Other brands give each business unit (or store) a fixed number of challenges in which they can ask for an additional shop. Managers responsible for that business unit can request a reshop for any reason. However, when the fixed number of disputes is exhausted they lose the ability to request a reshop. This approach is fair (each business unit has the same number of disputes), it reduces the administrative burden on a centralized arbiter, and reduces the potential for massive gaming of the system as there is a limited number of disputes.

Call to Action Analysis



A best practice in mystery shop design is to build in call to action elements designed to identify key sales and service behaviors which correlate to a desired customer experience outcome. This Key Driver Analysis determines the relationship between specific behaviors and a desired outcome. For most brands and industries, the desired outcomes are purchase intent or return intent (customer loyalty). This approach helps brands identify and reinforce sales and service behaviors which drive sales or loyalty – behaviors that matter.

Earlier we suggested anticipating the analysis in questionnaire design in a mystery shop best practice. Here is how the three main design elements discussed provide input into call to action analysis.

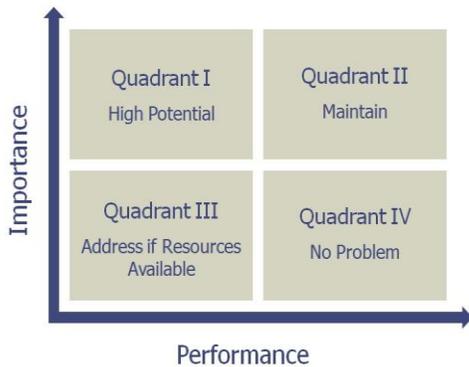
How

Shoppers are asked if they had been an actual customer, how the experience influenced their return intent. Cross-tabulating positive and negative return intent will identify how the responses of mystery shoppers who reported a positive influence on return intent vary from those who reported a negative influence. This yields a ranking of the importance of each behavior by the strength of its relationship to return intent.

Why

In addition, paired with this rating is a follow-up question asking, why the shopper rated their return intent as they did. The responses to this question are grouped and classified into similar themes, and cross-tabulated by the return intent rating described above. The result of this analysis produces a qualitative determination of what sales and service practices drive return intent.

What



The final step in the analysis is identifying which behaviors have the highest potential for ROI in terms of driving return intent. This is achieved by comparing the importance of each behavior (as defined above) and its performance (the frequency in which it is observed). Mapping this comparison in a quadrant chart, like the one to the left, provides a means for identifying behaviors with relatively high importance and low performance, which will yield the highest potential for ROI in terms of driving return intent.

This analysis helps brands focus training, coaching, incentives, and other motivational tools directly on the sales and service behaviors that will produce the largest return on investment – behaviors that matter.

Taking Action Part of Balanced Scorecard

A best practice in mystery shopping is to integrate customer experience metrics from both sides of the brand-customer interface as part of an incentive plan. The exact nature of the compensation plan should depend on broader company culture and objectives. In our experience, a best practice is a balanced score card approach which incorporates customer experience metrics along with financial, internal business processes (cycle time, productivity, employee satisfaction, etc.), as well as innovation and learning metrics.

Within these four broad categories of measurement, Kinēsis recommends managers select the specific metrics (such as ROI, mystery shop scores, customer satisfaction, and cycle time), which will best measure performance relative to company goals. Discipline should be used, however. Too many can be difficult to absorb. Rather, a few metrics of key significance to the organization should be collected and tracked in a balanced score card.

Coaching Best in class mystery shop programs identify employees in need of coaching. Event-triggered reports should identify employees who failed to perform targeted behaviors. For example, if it is important for a brand to track cross- and up-selling attempts in a mystery shop, a Coaching Report should be designed to flag any employees who failed to cross- or up-sell. Managers simply consult this report to identify which employees are in need of coaching with respect to these key behaviors – behaviors that matter.

Plan for Change

Such an informed feedback loop between customer surveys and mystery shopping will ensure the behaviors measured are aligned with customer expectations.

Finally, given mystery shopping measures employee behaviors against service standards, it is a best practice in mystery shopping to calibrate and align service standards with customer expectations. This is achieved by maintaining a feedback loop from customer expectations uncovered with surveys of customers back into updating both service standards based on these customer expectations and mystery shopping to measure and reinforce those standards. Such an informed feedback loop between customer surveys and mystery shopping will ensure the behaviors measured are aligned with customer expectations.

Even well-designed and administered best practices in mystery shopping research requires periodic adjustment. Performance scores eventually flatten out or cluster together, diminishing the value of the program as a tool for rewarding top performers and continuously improving quality. Periodic reviews should be worked into the program design so it can be kept relevant and useful, and so the bar can be repeatedly raised on service quality and employee performance.

Provider Selection

Truth be told...mystery shop data collection is largely a commodity, all mystery shop providers have access to the same pool of shoppers, and use similar technology to collect shop data. The source of differentiation is the extent to which a provider can help take meaningful action on the results.

Hire a provider that can be a partner. Large companies often employ an excruciating bidding process that rarely identifies the best vendor for their needs. They issue lengthy RFPs for mystery shopping that are meant to weed out the weakest contenders, but by asking bidders to commit to overly detailed and inappropriate specifications, they effectively eliminate more sophisticated companies at the same time. The typical RFP process creates an environment in which mystery shopping vendors over-promise in order to make the first cut, thus setting themselves up for failure if they win the account. In addition, it treats mystery shopping research as a commodity, regarding it as a bulk purchase of data rather than a high-value quality improvement tool. Companies have more success when they research the market carefully and identify the providers that have the knowledge and commitment to help them build a truly valuable program.

Conclusion

It is the employees who animate the brand, and it is imperative that employee sales and service behaviors be aligned with the brand promise. Actions speak louder than words. Brands spend millions of dollars on external messaging to define an emotional connection with the customer. However, when a customer perceives a disconnect between an employee representing the brand and external messaging, they almost certainly will experience brand ambiguity. The result severely undermines these investments, not only for the customer in question, but their entire social network. In today's increasingly connected world, one bad experience could be shared hundreds if not thousands of times over. Mystery shopping is an excellent tool to align sales and service behaviors to the brand.

Mystery shopping programs, when administered in accordance with certain mystery shopping best practices, identify the sales and service behaviors that matter most – those which drive purchase intent and customer loyalty.

Mystery shopping programs, when administered in accordance with certain mystery shopping best practices, identify the sales and service behaviors that matter most – those which drive purchase intent and customer loyalty.

For more information contact Eric Larse, co-founder of Seattle-based Kinesis, which helps companies plan and execute their customer experience strategies. Mr. Larse can be reached at elarse@kinesis-cem.com.