Aalto University
School of Business

THE USE OF MYSTERY SHOPPING RESULTS

From Retail Store Managers' Perspective

Master's Thesis

My Linh Tran

Aalto University School of Business

Management and International Business

Spring 2019

Author	My Linh Tran
Title of thesis	The use of mystery shopping results: from retail store managers' perspective
Degree	Master Degree
Degree programme	Management and International Business
Thesis advisor	Derin Kent
Year of approval 2019	**Number of pages** 64 **Language** English

Abstract

Since assessing service quality and performance in retailing has become more critical than ever, retailers are keen to use mystery shopping as a method to measure the customer service process. It is recognized that most companies adopt mystery shopping with the hope that performance measurement will encourage positive changes in managers and employees' behaviors, and subsequently, result in customers' satisfaction and better sales performance. However, transforming those information into accurate understanding and thus, leading to practical actions is not an easy task for many managers. Hence, this study aims to explore what store managers do with mystery shopping results and factors that influence their behaviors.

The theoretical part of this study reviews mystery shopping with a focus on the perspective of organizations adopting mystery shopping. The literature on management behaviors in using mystery shopping results is also presented. Following by a brief discussion of retail store operations and the role of retail store managers since the context for this study is retail stores.

The study uses a qualitative approach, and data was collected by semi-interview with seven store managers from the same retailer. The findings were analyzed using thematic content analysis and positioned within the process stages for performance management by Bourne et al. (2005), including data analysis, interpretation, communication, and taking actions.

The study found that store managers acknowledge the use of mystery shopping results for a variety of purposes. However, all respondents are much concerned about the reliability of mystery shopping results. Store managers then rely on their intuition and work experiences to analyze and interpret the results. It is implied that store managers only adopt the results on criteria that they found reasonable and keep in mind that the results only reflect a partial view of the store's operations. Besides, most respondents perceive getting a high mystery shopping score as a way to avoid the problem with their direct managers - area managers. However, store managers show no interest in getting the reward liked with mystery shopping as they see it unreachable. This study also found that store managers are sensitive about the well-being of sales staff, and thus, store managers try to avoid unnecessary stress for staff from the mystery shopping results. Despite many complaints about mystery shopping and its results, most store managers accept mystery shopping results and rarely feedback about the program to higher managers. This can be explained as store managers feel they lack support from the company, and they all aware of the service quality orientation in their organization.

Keywords Mystery shopping results, retail store manager, performance management

Contents

List of figures

List of tables

1. INTRODUCTION

1.1 Background

"What gets measured gets managed."

Peter Drucker

Companies are keen on making performance measurable to manage their business easier and better. However, understanding and measuring customers' experience still be a great challenge for many organizations, especially in the service field where iporates tangible and intangible elements. In the past, customer complaints and customer surveys were the main tools to assess employee performance and provided service; however, those data can only provide limited insights on the service standards (Wilson, 2001). Furthermore, when customers are disappointed with the products or service, they rarely exprss their dissatisfaction and easily switch companies due to bad experience (Hesselink & Van Der Wiele, 2003). Globalization, competitive pressures and well-informed customers have made traditional practices no longer appropriate for measuring service performance. This implies the need for other methods to measure the service quality and gain competitive advantage, and mystery shopping has risen to be a popular tool to evaluate frontline customer service (Myers, 2007).

Mystery shopping is a method used externally by market research companies or internally by companies to measure service quality, or to ensure the compliance with regulations, or to gather specific information about products and services (PamInCa, 2009). Trained mystery shoppers act as ordinary customers and perform pre-defined tasks such as purchasing a product, asking questions to evaluate the provided service, and then produce reports on their experiences in details (Finn & Kayandé, 1999; Wilson, 2001). Fronline staff who are evaluated by mystery shopping are not aware of their participation in the observation (ESOMAR, 2005). The organizations evaluated by

mystery shoppers will use the mystery shopping results as learning materials and take actions accordingly to improve their performances.

1.2 Research gap

According to the Mystery Shopping Providers Association, the value of the mystery shopping industry is 2 billion dollars worldwide (MSPA, 2018). Although mystery shopping has become a standard tool in the field, there is still few academic research about the value of this method (Wilson, 2001). The primary topics in recent research on this phenomenon include training mystery shoppers (Beck et al., 2004), the relationship between mystery shopping and service quality (Beck & Miao, 2003), the role of mystery shopping in service performance (Wilson, 2001). While the majority of current research focuses on the perception of mystery shoppers and sales agents – those directly involved in the mystery shopping process, insufficient attention has been paid to managers and how managers at different levels make use of mystery shopping results.

Few studies on the use of mystery shopping results are in line with the fact that there is much less literature on the use of performance measures, compared to research on performance measurement (Bourne et al., 2005, Ukko et al., 2007, Radnor & Barnes, 2007). Due to the sensitiveness of the performance information, companies are reluctant to reveal their practice of using performance measures (Bell, 2010), and thus, researchers have limited access to study the use of performance measures. Radnor & Barnes (2007) have called for further research from the operational management perspective towards performance management.

Furthermore, while mystery shopping has been studied in various contexts and from different theoretical perspectives, it remains that most mystery shopping literature has been quantitative.

1.3 Problem and Research questions

This study aims to address those mentioned gaps by exploring how store managers draw upon mystery shopping results in Vietnamese retail store context. Specifically, this study will explore, what store managers do with mystery shopping results and what are factors influencing their behaviors. According to Erstad (1998), mystery shopping should be implemented with both employees and customers in mind so that the outcomes can benefit not only customers but also employees. Store managers are frontline managers who often interacts with both customers and employees; hence, their insight into the use of mystery shopping results would be useful for better understanding of mystery shopping management.

As Wilson (2001) suggests the examination of the mystery shopping to consider contextual factors, this study opts to focus on the context of the retail stores in Vietnam. According to Dabholkar et al. (1996), the retailing context is different from other service settings due to its dual offering, aiming to provide both products and services to customers. Berry (1986, p.3) stresses the importance of service in retailing to the extent that "retail businesses are services businesses." In retail, service quality is a decisive factor, which determines customer loyalty in retail outlets (Sainy, 2010).

With a vast population, fast urbanization rate and a steadily increasing number of middle and high-income consumers, Vietnam is emerging as an essential retailing market. In the context of retail stores with numerous challenges such as cost-cutting measures, high employee-retention, and more savvy customers, mystery shopping is thus designed to address these challenges by evaluating and reporting the customer service at the frontline. However, it is still a challenge to produce objective evaluations from a third view since the outsiders cannot capture numerous factors that may influence the daily store operations. These challenges also require store managers with specific skills and efforts to maximize the use of mystery shopping results.

Mystery shopping can help managers by providing data for assessing service performance, coaching employees, and improving sales performance. However, in

reality, mystery shopping programs often suffer from poor design, unreliability data, lack of support from stakeholders. As a result, managers may have to deal with internal complaints and struggle to transform the results into accurate understanding and appropriate actions. Furthermore, Wilson (2001) claims that the relationship between mystery shopping results and customers' satisfaction and behaviors is not always clear. This implies that what managers do after receiving mystery shopping reports matters, and this study hopes to provide more insight into this phenomenon.

It is recognized that most companies make use of mystery shopping with the hope that performance measurement will encourage positive changes in managers and employees' behaviors, and subsequently, result in customers' satisfaction and better sales performance. By reading mystery shopping reports, store managers can receive early signals that frontline service quality and store operations are doing well, or there is some problem that needs to be improved. However, transforming those information into accurate understanding and thus, leading to practical actions is not an easy task for many managers (Franco & Bourne, 2003). Hence, this study further concerns with exploring factors that influence store managers' behaviors toward mystery shopping results.

The purpose of this study is to explore the use of mystery shopping results by store managers in retailing. Qualitative data from interviews will be analyzed to answer the following specific research questions:

1. What do store managers do with mystery shopping results?

2. What are factors influencing store managers' behaviors toward mystery shopping results?

Considering the widespread use and usefulness of mystery shopping, the findings from this study can provide valuable information for retailers who adopt the mystery shopping. Besides, researchers can use the results of this study as an empirical study for further qualitative and quantitative researches on behaviors towards using mystery

shopping results. In sum, a better understanding of the use of mystery shopping results is needed for both academic and practical interest.

1.4 Research approach

This study concentrates on the personal experiences of the store managers. This approach allows a clearer picture of store managers' perception regarding using mystery shopping results, which still lacks studies in the mystery shopping literature. Since the knowledge on the use of mystery shopping results is currently limited, qualitative research and interpretive perspective are appropriate for this study. Store managers in the same retailer would be interviewed based on an interview guide. Semi-structured interviews are piloted, and adjustments are made for better data collection.

1.5 Structure of the study

The introduction chapter summarizes the background of this study, following by defining the research gap, leading to the main research problem and research questions. Then the research approach and outline structure for the study are introduced.

Following the introduction chapter, the second chapter provides an overview of mystery shopping. First, by looking at the definition to understand the critical characteristics of this term. Secondly, the chapter discusses mystery shopping from the perspective of service quality measurement to understand what gets measured by mystery shopping. Further on, the chapter presents literature on management behaviors in using mystery shopping results. Next, a brief discussion of retail store operations and the role of retail store managers is presented since the context for this study is retail stores. Lastly, a theoretical framework is presented for data analyzing.

In the third chapter, the research methodology for this study is introduced. While chapter four presents the findings and discussions. The last chapter summarizes the findings and further present limitations and implications for management and further research.

2 LITERATURE REVIEW

This chapter first gives an overview of mystery shopping with a focus on the perspective of organizations adopting mystery shopping. I then review mystery shopping from the perspective of service quality measurement to understand service quality dimensions as what gets measured by mystery shoppers. Next, I present literature on management behaviors in using mystery shopping results. Also, a brief discussion of retail store operations and the role of retail store managers are offered since the context for this study is retail stores. Lastly, a theoretical framework is presented for data analyzing.

2.1 Mystery shopping

2.1.1 Definition

According to Wilson (1998), mystery shopping is a form of participant observation technique. Participant observation, origins in cultural anthropology, is a qualitative method where researchers interact and observe subjects. Through observations, researchers can understand those hard to explain via communication such as norms, cultures, attitude, and behaviors.

In general, mystery shopping is a tool to evaluate certain products, services and the overall experience of the everyday customer, to assess the compliance to regulation, or to improve the service standard (PamInCa, 2009). This technique can be performed externally by third parties such as mystery shopping providers or internally by companies themselves to achieve various purposes. The site evaluated by mystery shopper generally does not know about the mystery shopper's specific identity and his or her presence. Therefore, the observers would be called mystery shoppers or ghost shoppers.

Thus far, mystery shopping has been developed and widely used in different industries and most common in the service context, especially in those fields characterized by direct interaction between sellers and customers. Companies who use mystery shopping can range from the private to the public sector. If mystery shopping successfully practiced in the state and local government, it can help detect fraudulent behavior (Allison, 2009).

In the retail store context, the observers or mystery shoppers would act as ordinary customers and perform pre-defined tasks such as purchasing a product, asking questions and then provide detailed reports on their experiences in an objective way (Finn & Kayandé, 1999; Wilson, 2001; Turner, 2007). Furthermore, depends on the purpose of mystery shopping program, mystery shoppers can perform other tasks such as taking photographs, counting customers or inventory, recording the time to perform a specific task with a timer. In sum, mystery shopping is widely used as a tool for service quality measurement. The following section will review this particular aspect of mystery shopping in greater details.

2.1.2 Mystery shopping as a tool for service quality measurement

To understand mystery shopping as a tool for service quality measurement, it is critical to know why organizations seek to measure and manage service quality and adopt mystery shopping for service quality measurement. Further on, to understand what mystery shopping aims to measure and evaluate, the literature on key service quality dimensions will be presented, with a focus for the retail context.

Service quality measurement

According to Strawderman & Koubek (2008), service quality is the achievement of meeting customers' needs, wants, values, and expectations. Zeithaml et al., 1996 define service quality as an evaluation of the 'overall excellence of services from the customers' point of view.' Researchers generally agree that excellent service quality helps increase customers' satisfaction and loyalty. In today's competitive and consumer

demanding market, good marketing can help bring customers, but excellent customer service is vital to turn customers to be loyal (Zeithaml et al., 1996; Chadha & Kapoor, 2009). Promoting service quality also leads to profit generation (Parasuraman et al., 1988). Proven by both theory and practice, excellent service quality results in competitive advantages for companies (Zeithaml et al., 1996; Martinelli & Balboni, 2012). Hence, there is a constant need for measuring service quality to serve customers better and remain competitive in the market. This also explains why researches on service quality have dominated service and marketing literature since the 1980s.

Measuring service quality has always been a hot topic in business management (Parasuraman et al., 1988). As customers value service excellence, managers strive to define what important to customers and get dimensions for the service right. However, service quality is immaterial and hard to measure (Urban, 2013). Besides, service quality dimensions vary in different service sectors and contexts (Pollack, 2009). As a result, there is no consensus on the service quality measurement. However, there are still two main measurement paradigms in assessing service quality. One perspective measures the gap between customers' expectations and the perceptions of provided actual service, while the other focuses on evaluating the service performance from customers' perceptions.

In this study, the literature on mystery shopping would be based on service quality measurement and take the view of performance measurement. Precisely, mystery shopping measures the performance of service quality from customers' perceptions. This method does not focus on measuring the gap between customers' expectations and the perceptions of provided actual service.

Why organizations adopt mystery shopping as a tool for service quality measurement

According to Wilson (2001), the mystery shopping method outperforms traditional customer surveys by offering valuable measurement for service quality. A study by Finn & Kayandé (1999) regarding the psychometric quality of mystery shopping shows that mystery shopping data is more valid and reliable compared to data collected from

customer surveys. Based on single service visits, mystery shopping allows the evaluation to be more objective, while customer surveys are slightly biased since the opinions may be influenced by previous service encounters (Lowndes & Dawes, 2001).

Moreover, mystery shopping measures the whole process instead of the outcome of a service experience only. For example, customer surveys or customer complaints can only provide general opinions about the service based on the customers' memories. These methods cannot offer much detail information for organizations to identify mistakes and correct the wrong activities and procedures.

Service quality dimensions as what gets measured by mystery shopping

As a tool for service quality measurement, mystery shopping would measure some service quality dimensions to a certain extent. Some dimensions are less tangible and thus, identifying suitable measures for mystery shopping is a significant challenge. In this section, I will introduce three service quality dimensions in the retail context as what mystery shoppers measure and evaluate at a retail store. The three dimensions and their sub-dimensions will be discussed as below.

Physical aspects

Physical environment quality has first been studied by Kotler in 1973 for its environmental influences on consumer behavior. Physical aspects are defined by Elliott et al. (1992) as the physical features of the service process. Several researchers have found that physical environment quality is one of the essential aspects of customer evaluation of service quality (McDougall & Levesque, 1994; Rust & Oliver, 1994). The research conducted thus far suggested the following variables to compose the physical environment: the store's surroundings, the merchandise, the store's equipment, the comfort and the ambience, cleanliness, safety and security (Parasuraman et al., 1988; Dabholkar et al., 1996; Vazquez et al., 2001; Brady & Cronin, 2001; Sureshchandar et al., 2001).

In general, these variables concern either the presence, the quality or the appearance of physical factors within and around the store, and the comfort those factors provide for the customers. In terms of mystery shopping, physical aspects are less subject to change and quite easy to be observed.

Human interaction

Brady & Cronin (2001) define interaction quality as the interpersonal interface between service providers and customers during the service delivery process. The quality of human interaction is evaluated by variables such as body language, tone of voice, responsiveness, expertise, problem-solving skill (Dabholkar et al., 1996), and the employee is evaluated by, for example, friendliness, attitude, the level of enthusiasm, and appearance (Parasuraman et al., 1988; Dabholkar et al., 1996; Vazquez et al., 2001; Sureshchandar et al., 2001; Brady & Cronin, 2001).

Study of Gronroos (1982) stresses the importance of human interaction quality as the factor that has the most significant effect on the overall service quality perceptions of consumers. The human element, however, is hard to be measured and evaluated on a fair basis by mystery shoppers due to its intangible factors. Human is always hard to be measured, and the problem is more complicated when a human assesses a human.

Policies and Proficiencies

This dimension includes sub-dimensions concerning the policies of the service provider and its proficiencies, for example, compliances, administration, the processes and procedures (Parasuraman et al. 1988; Dabholkar et al., 1996; Vazquez et al., 2001; Sureshchandar et al., 2001; Brady & Cronin, 2001). Policies and proficiencies are essential dimensions since customers would always expect the service delivery processes to be standardized. Maintaining the standardized service is crucial for retailers; managers want to make sure that employees follow the policies and procedures to perform preferred behaviors, and thus, serve the customers better.

It is hard for a mystery shopper to interact with the shop-location once and evaluate the site's policies and proficiencies properly. Some contextual factors may affect the compliance with policies and procedures, that only service providers know those factors while mystery shoppers cannot understand. For example, a mystery shopper, who encounters a product that is out of stock, can easily assume that the store lacks proper administration. The mystery shopper, however, is not aware that problem may come from the supplier, and that is out of the control of the store.

2.1.3 The process to produce mystery shopping results

The following part will discuss the process to produce mystery shopping results. The process is presented from the view of organizations adopting mystery shopping, where managers can involve and be a part of the designed process.

Phase 1: Defining benchmark

Mystery shopping goals and criteria are built based on the organization's needs and purposes. When mystery shopping is performed by a watchdog company or market research company, the goals and criteria will be discussed and agreed between the mystery shopping service provider and the client company.

The criteria for evaluation are defined in the form of a checklist, and mystery shoppers can use the checklist as a guide for their observation. The checklist addresses a set of activities that are related to the specific workplace and job position. In general, the list covers overall quality service dimensions and key performance indicators that reflect company vision and mission (Zeithaml et al., 1990). According to Wagnerová & Baarová (2008), the pre-determined criteria in the checklist indicate preferred behaviors among employees by the management board.

Phase 2: Data gathering

During the service assessment, a mystery shopper will behave as a regular customer to observe the performance of critical services such as the store environment, the level of hospitality, product availability. The mystery shopper may also secretly take photos and use photos as evidence for their feedback. The qualified mystery shopper must be trained about the checklist, specific situations, and the context of the shop-location to provide an independent, objective, and critical evaluation.

Phase 3: Reporting

After the encounter, mystery shoppers will recall what they have seen during the visit to fill in a checklist and document the mystery shopping report. Rubel (1995) states that the report needs to address several audiences who need different types of information. The mystery shopping results should not go to top management but rather to the people directly involved, including the frontline employees (Spooner, 1985). Once the results have been published, reviewed and improvements have been made, then the top management should receive mystery shopping reports. Besides, communication is crucial in this phase, as Burnside (1994) states that mystery shopping results should be positively delivered to employees. Furthermore, managers of the site evaluated by mystery shoppers should feedback to mystery shoppers about the quality of the reports (Burnside, 1994).

Phase 4: Value acquirement

Managers of organizations evaluated by mystery shoppers will use mystery shopping results to identify any change needed to improve the service quality. Besides the use as learning materials, the results can be linked to bonuses, awards, and prizes (Dorman, 1994). In general, according to Morrall (1994), managers must be informed of the value has been achieved from the mystery shopping program. However, both theory and practice still have not paid much attention to the value acquirement phase of the mystery shopping process, and thus, this study aims to contribute to this gap. The following

section will provide more information on what managers do after receiving mystery shopping reports.
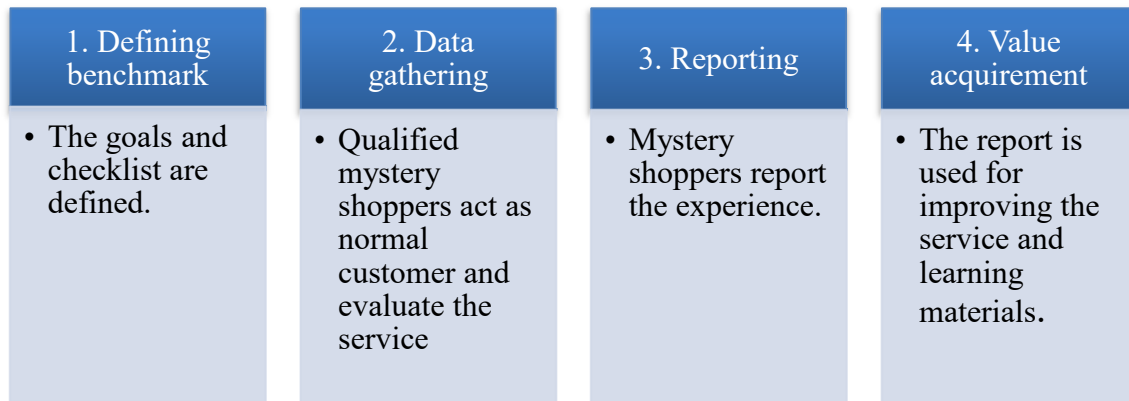
1. Defining benchmark	2. Data gathering	3. Reporting	4. Value acquirement
• The goals and checklist are defined.	• Qualified mystery shoppers act as normal customer and evaluate the service	• Mystery shoppers report the experience.	• The report is used for improving the service and learning materials.

Figure 1: Mystery shopping process

2.2 Management behaviors in using mystery shopping results

As stated above, there is a limited focus on the literature of what managers do with mystery shopping information. Furthermore, the current literature has not addressed the use of mystery shopping by different level of managers in an organization. However, the following part will review related literature on this topic. I first review the existing studies on the use of mystery shopping results. Then I will present the literature on several factors that may influence managers' behaviors toward using mystery shopping results.

2.2.1 The use of mystery shopping results

Companies strive to make good use of mystery shopper results since a vast amount of money and resources have been spent in the mystery shopping program. This part will present the literature on how managers use mystery shopping results.

Analyze and communicate the results

According to Cramp (1994), managers should analyze mystery shopping results with the record of previous visits in mind. Since a single mystery shopping visit cannot review

performance for an employee or a store. Moreover, to increase the validity and reliability of mystery shopping results, managers can cross-check mystery shopping results with results from other sources such as customer satisfaction surveys or customer complaints (Restaurant Business, 1989).

Communicating the results to employees is recognized as a crucial part of mystery shopping management. According to Buchner (2007), control theory focuses on feedback as a means to shape behavior. As the employees receive feedback from mystery shoppers on their behaviors, they become more aware of the discrepancy between what they are doing and what they are expected to do, and thus, the staff can take corrective actions to minimize the difference. Managers thus have the responsibility to find the most suitable form to communicate the results to the employees, and it should be in a positive way (Burnside, 1994). According to Parker (1988), managers should feedback the mystery shopping results to employees as soon as possible for best results. The immediate feedback helps capture staff' attention and motivate staff for further improvements since the frontline staff can see what behaviors can be rewarded. At Burger King, mystery shopping results are published on bulletin boards at the restaurant for all employees to view (Burnside, 1994). Some retailers not only publish the mystery shopping results at their single outlet but also nationally (Cramp, 1994).

Another way to communicate mystery shopping results calls for depersonalizing the results by not identifying the individual staff in the report and avoiding follow-up victimization. Cramp (1994) states that employees most concerned about how mystery shopping results are used. If an employee becomes worried that mystery shopping results may make his or her lose the job, the staff then will not focus on serving the customer (Dorman, 1994).

Mystery shopping results and actionable recommendations

To improve service quality

According to Wilson (2001), organizations can use mystery shopping results to identify weak points in the service delivery, to motivate service staff for further improvement, or to benchmark and monitor the service standards across channels of retails. It is also proposed that managers can use mystery shopping results in addition to results from other methods to measure customer satisfaction (Hesselink & Van de Wiele, 2003, Van der Wiele et al., 2005).

In terms of a diagnostic tool, mystery shopping can provide feedback on important services and help managers identify where to improve. Managers then can allocate reasonable capital, technical, and human resources to maintain or enhance the service level (Wilson, 2001). The customer insight can be different from the viewpoint of the company and thus, is essential to be considered for better decisions. Furthermore, with the regular evaluation of mystery shopping results, managers can monitor and compare the service standard at a different time such as peak, weekend, high season and non-peak, weekdays, low seasons.

Mystery shopping not only helps managers in detecting what needs to be improved, but it also encourages employees to follow up the code of conduct and work better as the staff knows they are evaluated by customers (Wilson, 2001). The Hawthorne effect has acknowledged the tendency of some people to work harder and perform better when they are working under observation. However, many researchers still argue whether the group of people made improvement due to various factors which influenced their behaviors or simply because of the attention they had (Kompier, 2006).

Furthermore, mystery shopping is a useful tool to monitor the service standard across retail channels. Finn & Kayandé (1999) argue that mystery shopping allows the assessing of whole branches, rather than only one service facility. Managers can use mystery shopping to assess whether customers are being treated equally across the channel (Morrall, 1994). Most mystery shopping practices often provide the evaluation

score based on a pre-determined scale. By reading mystery shopping reports, managers can check the overall score of the whole organization as well as each service facility with the highest and lowest score. In a way, the result of mystery shopping can help managers to develop a healthy competitive environment among subsidiaries and employees.

However, Wilson (1998) argues that the impact of mystery shopping to the improvement of service performance only lasts for the short term since little effort has been made to prolong the impact of mystery shopping results. Mystery shopping is considered as a part of company's performance management, and the idea is to learn from customers' point of view which aspects of service quality are most crucial and where need improvement to take further actions. However, according to Wilson (1998), there are few attempts to integrate mystery shopping results with other variables of service delivery, such as customer satisfaction, staff attitudes, and sales figure. As a result, the link between mystery shopping results and customers' satisfaction and behaviors is not always evident and cannot meet the expectations of the management board.

Link performance to a reward system

Mystery shopping can be seen as a part of a behavior modification program (Dorman, 1994). To change employee behavior, the result of mystery shopping is often linked to bonuses, awards, prizes, or other incentives in the organization. According to Cramp (1994), 60 percent of retailers link mystery shopping results to their reward systems. Reward and recognition can be designed on an individual or team basis (Dorman, 1994), either in monetary or non-monetary forms, and can be quarterly, annually, or linked to individual shopping visits. Frontline staff are representatives of the organization and are first ones can identify any potential bottlenecks. Hence, companies can use mystery shopping to identify outstanding staff with excellent service skills and reward them for their excellent work.

Rewards, incentives, and bonuses are believed to reinforce positive behavior among staff. However, for most managers, to ensure reward with cognitive self-evaluation and fairness considerations is not an easy task. Regarding mystery shopping, it is believed that this method can provide objective evaluations from a third party; however there are still arguments whether the outsiders can capture numerous factors that may influence the daily store operations and thus, offer objective evaluations. As a result, many organizations and researchers still concern about the reliability and fairness of mystery shopping results.

Training the employees

Managers can use the feedback from mystery shoppers as a necessary training tool since it can be used to identify what skills need to be developed, and thus, enhance training effectiveness (Spooner, 1985). Base on mystery shopping results, managers can compare training and levels of customer service before and after training. Parker (1988) claims that training and rewards linked with mystery shopping can help improve up to 20 percent of overall customer service. The pre-determined criteria in mystery shopping checklist also indicate favored behaviors for employees from the management board (Wagnerová & Baarová, 2008). Hence, mystery shopping allows managers to determine if the frontline employees are providing appropriate services. Sobel & Hines (1990) stress the importance of both technical training regarding job functions and behavioral skills training for frontline employees. The training success can be reviewed during later mystery shopping visits (Morrall, 1994).

According to Cramp (1994), mystery shopping provides useful managerial implications and actionable recommendations. The underpinning logic here is that since the mystery shopping program requires a considerable amount of resources and finance, retailers would try to make use of mystery shopping results. "If the action does not follow measurement ... most of the efforts ... will be wasted" (Franco & Bourne, 2003, p.15). Organizations adopt mystery shopping as well as other performance measurements partly with the hope that these measures will influence the behavior of managers and

subsequently, result in positive performance. However, many managers still find it difficult to accurately analyze and interpret the results to understanding that 'helps to inform effective actions' (Franco & Bourne, 2003).

Furthermore, the influence on managers' behaviors from mystery shopping information may not be entirely positive. There are still negative implications of performance measurement. The following part will present a brief review of how performance measures may potentially negative affect managers' behaviors. This will help provide a comprehensive understanding of what store managers can do with mystery shopping results.

Potential negative influences on managers' behaviors

According to Smith (1995), performance measurement may entail those following unintended consequences.

- Tunnel vision: Managers focus only on areas that being evaluated by performance measurement system and exclude other essential areas that not captured by performance indicators. Especially if managers are given incentive for reaching the target reflected by the performance measurement, managers can ignore those aspects not measured and rewarded.

- Sub-optimization: Managers tend to pursue their narrow objectives rather than engaging in the whole organization's goals.

- Myopia: Managers may focus on short-term issues at the expense of long-term outcomes. This is a common issue to many measurement systems, and managerial myopia arises when the time horizon of the manager mismatch the time horizon of the project he is responsible. Especially when it often takes a long time to yield a valued outcome.

- Convergence: Managers tend to stay within recognized limits. They have no motivation to be outstanding.

- Ossification: Managers tend to stick to old familiar measurement rather than adopt a new or innovative experiment.

- Gaming: Managers can manipulate behavior to generate another outcome. For example, managers might intently underperform to avoid being set more demanding targets. This behavior is often seen when dealing with financial objectives.

- Misrepresentation: This behavior includes generating false indications, fraud, and creative accounting.

In general, the influence of measures and performance information on manager behaviors are diverse, ranging from positive to adverse effects. The presented literature has addressed how managers can use mystery shopping for positive purposes, as well as briefly review possible distortions arising from the use of performance measures.

As Ghosh & Lusch (2000) state that there are a 'plethora' of variables to contend with in retailing context, several factors can interfere with manager's decision-making and thus how managers make decisions upon mystery shopping results remains unclear. The presented literature review has justified a study that seeks to understand better how managers use mystery shopping results in the context of retail operations.

2.2.2 Factors that influence managers' behaviors

Numerous academic researches have recognized a broad range of internal and external factors that impact on manager behaviors towards performance measures. Since the external factors are beyond the control of the organization, this study would focus on internal factors that managers can aware and manage to a certain extent. This section will provide a brief review of factors that may influence managers' behaviors toward using mystery shopping results.

As mentioned before, the logic behind adopting performance measures is that managers can use the gathered information to guide decision-making and actions that lead to

stronger performance outcomes in the future (Neely et al., 1995; Waal, 2003, Franco-Santos & Bourne, 2005). Those researchers imply the link between performance indicators and managers' behavior relies on managers' perceptions and interpretations. Also, it is recognized the potential of various factors that may influence managers' behaviors. The following section will discuss the factors involved in the use of mystery shopping results.

Managers' perceptions of mystery shopping

It is critical to understand managers' perceptions since a positive view towards the valid and benefit of a performance measure; in this case, mystery shopping results can direct managers to desired behaviors. On the contrary, an unfavorable perception will weaken the effect of performance measures on managers' behaviors.

According to Webb (2004), managers more involved in performance measurement when the performance measurement system is perceived as reasonable. Moreover, managers tend to ignore "inappropriate" performance indicators (Bourne et al., 2005). Managers have their perceptions and can decide which is impossible, and consequently, they do not attempt to reach the "inappropriate" target. In this case, performance measures do not influence behavior. For example, if a measure indicator is perceived as beyond the control of the manager or is unfair in some way, then this might have a negative impact on the influence. On the other hand, favorable perceptions of measures can motivate managers to perform positive behavior. According to a study at the store level of a UK food retail by Wilson (2000), inappropriate performance measures are found to be the reason for the managers' inability to influence outcomes.

Morrison et al. (1997) state that most organizations are still concerned about the human factors of mystery shopping. Precisely, managers are concerned about the reliability of the mystery shopping results and reliability of mystery shopper who gathers data. Hesselink & Van der Wiele (2003) state that the valid of mystery shopping data depends on the person gather data – mystery shopper and reliable data must be reflected through the whole designed process from gathering to the reporting process.

Schwartz & Schwartz (1955) state that the observation by mystery shoppers should reflect reality to gain reliable data. However, as mystery shoppers have to remember what happened during the visit to fill in the feedback form, some mistakes may occur due to memory failures. For example, with a complex checklist that takes time to remember all criteria, the mystery shopper may forget or skip one or more requirement on the list. Any missing data can threaten the accuracy of the report, and thus affect the reliability of the mystery shopping results. Besides, other physical or psychological factors associated with mystery shoppers may affect their perceptions and evaluation. Fatigue, for example, can influence the quality of their assessment, especially when a significant amount of data and careful observations are required (Guerrien et al., 1993).

Besides, in some situations, the objectiveness of mystery shopping observations is threatened by several cognitive factors and thus, may negatively affect the result of mystery shopping (Calvert, 2005). The mystery shopping experience is influenced by the behavior and interaction of both participants – mystery shopper and the frontline staff. This implies the differences among mystery shoppers such as age, gender can be reflected in mystery shopping results. For instance, female mystery shoppers are found to provide more accurate data than male, while younger adults are considered to be most reliable (Morrison et al., 1997). Morrison et al. (1997) also find out that male mystery shoppers tend to get priority over women in department stores. This means that male mystery shopper may give more favorable evaluations than his female counterpart since he can receive better customer service in the department store.

Last but not least, there are several concerns regarding the generalizability of mystery shopping data. Traditionally, the sample size is a crucial factor that determines the reliability of a method. Can gathered data from an encounter reliable enough to be used for the evaluation of service quality? Turner (2007) argues that each observation is a valid "snapshot" of the service experience, rather than a representation for the all "population" of such experiences.

Other factors

Besides the managers' perceptions, Morgan et al. 's (2005) found that the "customer orientation" of an organization is an influential factor. The stronger customer orientation of a business, the more likely that company makes greater use of customer related information. Furthermore, training related to the use of performance indicators and the timeliness of reporting has found to influence performance measurement effectiveness (Wilson, 2000). Another study on performance management practice by Ukko et al. (2007) have identified the maturity of the measurement, reward system, and education levels of the employees as influencing factors (Ukko et al., 2007).

It is appropriate now to acknowledge several factors that may influence managers' behaviors and performance measurement effectiveness. Moreover, each different context will have different factors. The following part will give an overview of retail store operations and the role of store managers as the context for this study.

2.3 Context for the study

2.3.1 Retail store operations

Retailing is "the management of resources to supply the product and service needs of the end consumer, encompassing the supply chain of any physical products and the exchange processes involved." (Davies, 1993, p.6). Retail operations include a broad spectrum of activities to maintain the store function well on a daily basis. The key retail-specific activities include sales management, merchandising replenishment, inventory management, shrinkage control, customer service, staff management.

Retail is a competitive environment where there is constant pressure to maximize profitability and minimize costs. Gagnon et al. (2005) assert that there are a variety of internal constraints that reduce the stores' capabilities to operate efficiently, for example, cost-cutting measures, poorly trained employees, or ineffective management. Such practices have made the retail workforce mainly consisting of women and students

(Mason & Osborne, 2008). These frontline staffs have to be proactive and flexible on the sales-floor to meet varying customer demands (Broadbridge, 2002).

2.3.2 Role of store managers

Considering a wide range of operational in-store activities, the role of a store manager is a challenging role. Store managers have tons of responsibility to ensure the efficient and effective delivery of retail service delivery. Their performance is monitored and evaluated against various KPIs. In terms of financial pressures, store managers are expected to maximize sales and minimize costs and shrinkage. In terms of non-financial aspects, store managers have responsibilities for housekeeping, compliance with company policies and procedures, managing sales staff, and maintaining a high level of service quality. Mystery shopping KPI is one of those non-financial KPIs.

Most frontline staffs in retailing are low-paid and part-time. Hence, they are easy to switch to another company that offers a higher salary or a better job. As a result, the employee turnover in retailing is always high. Thus, there is a considerable pressure for managers to keep and motivate staffs to get the job done or to accept the fast pace with constant changes, especially when there is resistance from staff (Varley & Rafiq, 2004).

Store managers report to area managers, or operation manager, and have to follow broad organizational policies or directions from their boss. In a way, store managers' daily work tasks are heavily prescribed by the organization. However, in the store context, the store manager is the boss and is responsible for all aspects of the store performance. Despite various operational activities that allow the autonomy to a certain extent, the centralization of business processes, tight budgets, and organizational policies has reduced the flexibility and choices of many retail store managers (Grugulis et al., 2010).

2.4 Theoretical framework

As the use of measures is often investigated through the stages in the underlying process, this study will analyze what store managers do after receiving mystery shopping results through stages, including:

1. Data analysis: Analyze the data to turn the results into performance information (Lynch & Cross 1991; Neely, 1998)

2. Interpretation: Interpret the information to managerial implications (Simons, 1991; Neely, 1998)

3. Communication: Communicate the information to related people who will use the information for their decision-making (Bititci et al. 1997; Neely, 1998; Otley, 1999; Forza & Salvador 2000; Kerssens-van & Fisscher, 2003)

4. Taking action: Use the information in deciding what actions are needed to improve performance (Flamholtz, 1985; Simons, 1991).

This study adopts the following process for analyzing data on what managers with mystery shopping results.

Figure 2: Proposed process in using mystery shopping results (Adapted from Bourne et al., 2005)

Besides, the following framework will be used as a lens on factors influencing store managers' behaviors. It is developed from the proposed process in using mystery shopping results and based on the relationship between mystery shopping results and performance. The idea is that store managers use mystery shopping results to guide their

decision-making and behaviors with the hope that leads to stronger performance outcome in the future. However, several factors may interfere with the link between mystery shopping results and managers' behaviors, and thus, influence the outcome.
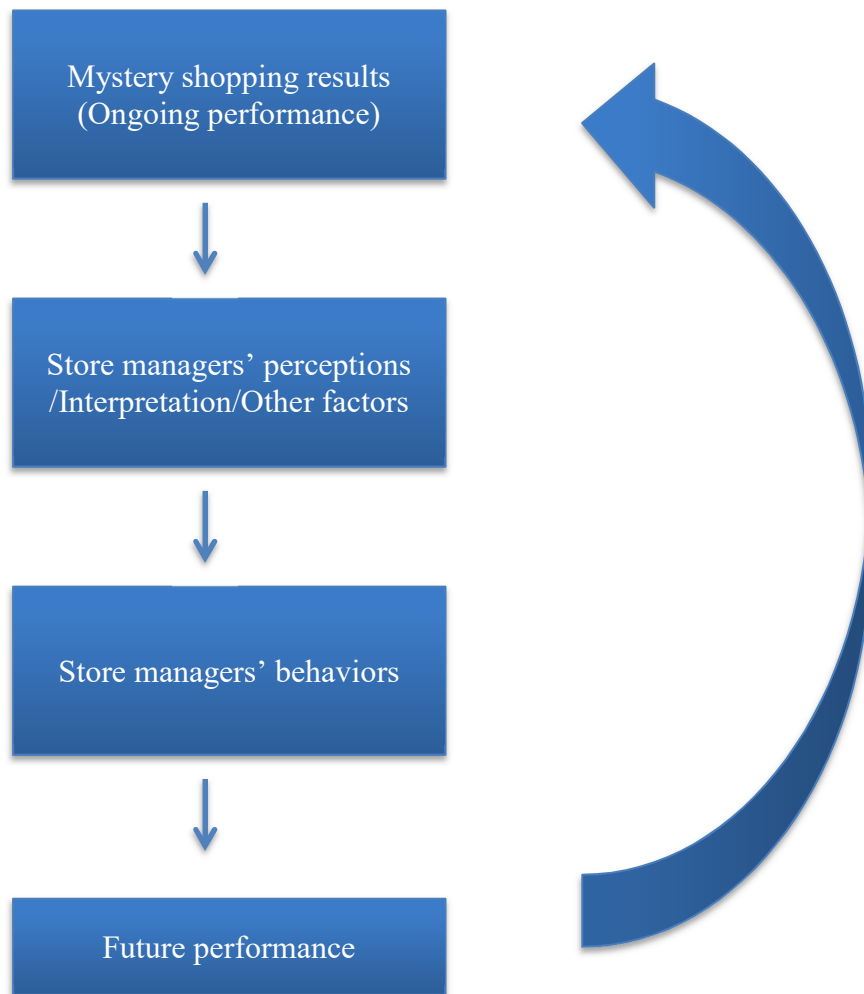
Figure 3: Proposed framework for analyzing factors influencing managers' behaviors

Summarize of literature review

The literature review has identified several theoretical gaps that this study aims to contribute. There are few literatures on mystery shopping, and most focus on the

reliability of this measurement. This is in line with the phenomenon that a majority of studies focus on performance measurement rather than performance management.

In this study, the literature on mystery shopping would be based on service quality measurement and take the view of performance measurement since mystery shopping measures the performance of service quality from customers' perceptions. The literature also presented key service quality dimension as what mystery shopping evaluates and measures. The context of retailing stores consists of various tangible and less tangible elements. That complexity makes the job of mystery shopping is more challenging than ever and requires managers specific skills and effort to maximize the use of mystery shopping results.

Despite limited literature on the management of mystery shopping results, this chapter has presented related studies on this topic. More specifically, measures are believed to be a useful tool to drive performance, and a study by Wilson (2001) has found that mystery shopping results can help organizations in various ways. However, in reality, the influence of measures and performance information on manager behaviors maybe diverse, ranging from positive to adverse effects. The underpinning logic here is that store managers use mystery shopping results to guide their decision-making and behaviors with the hope that leads to stronger performance outcome in the future. However, several factors may interfere with the link between mystery shopping results and managers' behaviors, and thus, influence the outcome.

This chapter also provided a brief review of retail store operations and the role of store managers to facilitate the understanding of the context of mystery shopping. Last but not least, proposed theoretical frameworks were presented for data analyzing.

3 METHODOLOGY

This chapter explains how research is conducted in this study. Starting with reflecting on philosophical underpinning including ontology and epistemology, and further explaining the decision on explorative, qualitative research. Then I present what and

how data was collected as well as how the analysis process was executed. The ethical concerns and evaluation of the study are also presented

3.1 Ontology

As ontology seeks to answer the question "What is the world?", ontology concerns the "real nature and the existence of people, the world as well as the relationship between those" (Eriksson & Kovalainen, 2008, p.13). This study takes the constructionism and interpretivism view, which assumes that social actors establish reality through their social interactions such as language and shared meaning (Eriksson & Kovalainen, 2008). In particular, this study assumes that social actors, including store managers, HR managers, sales staff, mystery shoppers, and other stakeholders, would interact and influence the perceptions and actions of each other. Through making interactions, store managers form their perceptions of using mystery shopping results, and store managers can tell what the reality is from their point of view.

Although retailers engage in using mystery shopping to control and improve their operations and service quality, store managers might have their way of using the results, and that may not follow the expectation of the company. There is no initial presumption in this study, except that factors are affecting the store managers' behaviors in using mystery shopping results.

From my own experience and by interacting with stores managers, I also participate in creating social reality. Also, if I interact with respondents at different times, the answers or results would be slightly different as well. Hence, information for this study would reflect the reality with the awareness that information has been interpreted through social conditions of the interviewees and mine.

3.2 Epistemology

Epistemology focuses on the question "What is knowledge and what are the sources and limits of knowledge" (Eriksson & Kovalainen, 2008, p.14). According to Saunders et

al. (2009), epistemology refers to acceptable knowledge in a field of study, and thus focuses on the researcher's opinions and perspectives. This study adopts subjectivist epistemology, which suggests that knowledge is available through social actors (Eriksson & Kovalainen, 2008).

In this research, I am interested in the perceptions and behaviors of store managers. I believe that store managers can tell stories that reflect their reality, their knowledge, and experiences in using mystery shopping results in retail stores in Vietnam. The context of retail stores in Vietnam is unique, and thus, only those store managers can provide information and understanding that may not be applied somewhere else in the world. Moreover, despite all stores operate under one retail chain, each store would have its unique context; hence, each store manager would have her perceptions of using mystery shopping results in her store context.

3.3 Research approach

This study adopts a qualitative research approach to gain an in-depth understanding of how store managers use mystery shopping results as well as factors that influence store managers' behaviors. I choose to conduct a qualitative study since it allows for more exploratory research rather than empirical testing (Marschan-Piekkari & Welch, 2004). Besides, the qualitative study allows representing the perspective and opinions of the respondents in the study (Yin, 2011), and thus, it is a suitable choice for this research. Furthermore, a qualitative approach focuses more on words rather than numbers and emphasizes a specific context (Maxwell, 2005). This can help identify contextual influencing factors that may affect the decision-making of store managers.

The research practice is more focus on measurement itself rather than on measurement management. As a result, the current insight on how mystery shopping results are used is currently limited. It is argued here that a qualitative approach to seeking insight and understanding would be most appropriate. Also, an interpretive study has much to offer in terms of understanding what managers do with mystery shopping results.

Qualitative research was conducted in the form of seven semi-structured interviews with retail store managers. Considering the research topic, my time and resources constraints for a master thesis, seven interviews for this study is reasonable to serve for replications, contrasts, and extensions to the phenomenon.

This study seeks information about how store managers make use of mystery shopping results and factors that influence their behaviors. Hence, semi-structured interview with open-ended questions would be useful to ask managers about their perceptions or ask them to give stories about their experiences with the mystery shopping program. Open questions allow respondents to choose a more flexible way to answer and provide more detailed responses (Eriksson & Kovalainen, 2008). Besides the interviews, I use secondary data such as website, mystery shopping reports, and any recommended materials by interviewees.

3.4 Research design

According to Yin (2011), the case study design includes five components as below:

1. The research question(s)

2. Research propositions

3. Unit of analysis

4. Data are linked to the propositions

5. Criteria to interpret the findings

In this study, the research question focuses on what retail store managers do with mystery shopping results and what factors influence their behaviors. As the research is explorative, no tangible propositions are formed at the beginning of the study except the presumption that factors may affect how store managers perceive and use mystery shopping results. The unit of analysis is the individual store managers from the same

retailer. Data are linked to the proposition through pattern matching and explanation building.

Adopted an interpretive perspective for this study, it is crucial to understand the context in which store managers operate and use mystery shopping results. Furthermore, information on how managers use performance results is quite sensitive, and thus, gaining access to the subject can be a significant challenge. Hence, the Health and Beauty retail chain Guardian Vietnam is chosen as the context for data collection since I have worked as a Management Trainee in this company for two years, from 2014 to 2016.

Guardian Vietnam is a Health and Beauty retailer that specializes in offering cosmetics, skin care, personal care, and health care. The company's target customers are Vietnamese consumers and the middle class. As a former management trainee at Guardian, I have worked as a store manager for six months and equipped myself with some understanding of the mystery shopping program which is used to evaluate the service quality in Guardian. This background also helps me to identify some store managers for this study and gain some trust from respondents.

The retail industry is changing dramatically, and Guardian Vietnam is not an exception. I have left Guardian Vietnam for about three years. Hence, to gain updated insights on the context of mystery shopping at Guardian Vietnam, I decided to have some pilot interviews to familiarize myself with the interview context. For the pilot interviews, I aim to gain information on strategies and policies that manage mystery shopping, the mystery shopping process and mystery shopping report, as well as how store managers use the mystery shopping results. Three store managers were then interviewed, and at the end of pilot interviews, I would ask store managers for any information that I can use for my study. Subsequently, I had a look at the mystery shopping report and store code of conduct. Then based on the results of pilot interviews and feedback from my supervisor, few adjustments were made for better data collection.

Before this study, I had no prior experience of a semi-structured qualitative interview. During the research process, I have become more aware of the need to balance between "active listening" and awareness of any preconceptions that blind the researcher to the perspective of the respondent (McCracken, 1988). Having listened through each interview recording and reflected on my performance as a researcher, I have tried to made adjustment to both the interview guide and the way I interview for better data collection. The following part will provide greater details on data collection and data analysis for this study.

3.5 Data collection

The purposive and snowball technique was used for sampling in this study. To target key informants, the purposive approach was initially applied to identify four participants from my network. Since the sample needs to reflect the qualifications, experiences, and knowledge of the research topic, respondents for this study have to be store managers and experience with mystery shopping program. The potential interviewees were contacted by visiting their working place or via Facebook. The research introduction and a description of interviewing (e,g. duration, language, recording) were mentioned in the invitation. Then four participants were later asked to recommend other potential participants from their networks who meet the research criteria.

In sum, I made seven semi-structured interviews with store managers. This small sample size allows me to "learn a great deal about the particular respondent and their response to the research questions" (Smith et al., 2009, p. 51). In line with standard practice, all respondents have been anonymized in this study. The following table summarizes key information of participants in this study.

Table 1: Respondent information

Case / Respondent	Gender	Birth year	Roles in Guardian	Years of Store Manager experience (Guardian)
A	Female	1985	Store Manager	7
B	Female	1990	Sales staff, Assistant Store Manager, Store Manager	3
C	Female	1990	Sales staff, Assistant Store Manager, Store Manager	2
D	Female	1992	Assistant Store Manager, Store Manager	2
E	Female	1987	Sales staff, Assistant Store Manager, Store Manager	1
F	Female	1991	Store Manager	2
G	Female	1990	Store Manager	1

The participant can decide time and location for the interview for her convenience. Besides, the place should provide a quiet and safe environment for the respondents to comfortably share his or her opinions, experiences. All interviews were in Vietnamese and lasted approximately 45 minutes. I asked the recording permission for accuracy and transcribing, and used my cellphone to record the interview.

Interviews are conducted in the Vietnamese language. Since all participants and I are native Vietnamese speakers, and I have been aware that some store managers are not fluent in English. Marschan-Piekkari & Reis (2004) argue that linguistic equality between researcher and participants produces the best interview outcomes. As

participants can use their mother language in the interviews, it helps acquire additional culture-related information, clarify potential misunderstandings, and benefits the study in many ways.

This study adopts a semi-structured interview since it allows more flexibility to ask questions in a convenient order and still get in-depth data. According to Smith et al. (2009), an interview guide would be useful to facilitate comfortable interaction and keep track of the covered topics in the interview. Hence, I made an interview guide that includes eight open-ended questions and some prompts relating to them, as shown in the appendix. Essential questions were developed to address the use of mystery shopping results and influencing factors on store managers' behaviors. Store managers were asked about the role of mystery shopping in their job and how they use the results for what purposes.

The interview should base on my interview guide; however, the questions were flexible and depended on the interviewee's answer. I would ask some theory-driven questions but keep in mind that those answers are at interpretative level (Smith et al., 2009) and would try to avoid leading and closed questions by staring my question more with "how", or "can you tell me about". During the interview, I would use prompts and probes to get more follow-up answers (Smith et al., 2009). Besides, I would pay attention to the order of questions to facilitate the information. Patton (2002) suggested that opinions and feelings questions should be asked before the knowledge questions as it is easier to understand the experience, then get to know more about what fact the interviewee understand. At the end of the interview, I would ask for any data that the store managers can provide for further reference and if there is any expectation from them for this research.

3.6 Data analyzing

Each interview was transcribed using Otranscribe website on my laptop. I transcribed word by word but omit some repeated words when respondents would clutter. I also edited and removed some words that I feel redundant for analyzing such as "uhm," "ahm" and phrases when people use when they do not know how to express their ideas such as "how do you say." As respondents and I speak Vietnamese as our first language, I decided to transcribe interviews in Vietnamese and analyze in Vietnamese as well. Then the emerging themes were translated into English as well as the quotes shown in the findings chapter.

Braun & Clarke's (2006) six-step of thematic analysis was adopted for this study. Each interview was analyzed individually, then followed by a cross-case analysis. The within-case approach allows capturing specific themes and issues, while the cross-case approach facilitates comparison between cases, as well as identify similar patterns or differences.

The overview of the 6-step framework by Braun & Clarke (2006) and how I conducted the analysis is presented below:

Step 1: Familiarizing with the data

In this step, I have read and reread the interview transcripts and other supporting materials until I feel familiar with the topic and can locate parts with rich information. After reading many times, I jot down my initial impressions and notes on each transcription, which helps me have an overlook of the interview content and not forget the important points.

Step 2: Initial coding

In this step, I start organizing my data and generate initial codes for each data group that relevant to the research question. I note down the actions and some factors and the meaning of those things from respondents' perspectives. I also pay attention to the language used and the context of the respondent to better interpret the findings. I

underline, note and highlight on the hard copies of the transcript to develop and modify the codes.

Step 3: Themes searching

According to Braun & Clarke (2006), a theme is characterized by its significance, but there are no fix rules for identifying a theme. In this step, I organize my codes to themes that answer the research questions and base on the theoretical frameworks. Braun & Clarke (2006) have categorized two theme levels, including semantic and latent themes. While semantic themes capture the "surface meanings of the data," the latent themes would look beyond what has been said by the interviewees and "examine the underlying ideas, assumptions, and conceptualizations" that shape the semantic data. (Braun & Clarke, 2006, p. 84). I listed all themes on a paper to easy view and modify later. Besides critical themes, I can have some 'miscellaneous' themes to group codes that I cannot categorize to a specific theme.

Step 4: Reviewing themes

At this stage, I review the themes and theoretical frameworks to see if the themes make sense and have enough supporting data for the themes. In case of any incoherence among themes, I would need to modify themes and revise again.

Step 5: Defining themes

According to Braun & Clarke (2006, p. 92), defining themes is to identify "what is the essence of each theme" and the relationship between themes. Also, if there are any subthemes, I need to define how the subthemes interact with each other and relate to the main themes.

Step 6: Writing

3.7 Ethical Concerns

Ethical concerns are not only reflected in this part of the study but would be considered throughout the thesis process from the research design to data collection, analyzing and reporting, as suggested by Saunders et al. (2009).

According to Saunders et al. (2009) and Eriksson & Kovalainen (2008), it is crucial to provide a clear account of purpose. Hence, when contacting potential participants, the purpose of this study was explicit. I provided interviewees with all necessary information, such as the interview purpose, my background, and a brief research introduction. I also explained what I would like to interview and upon agreement, an interview was set up. Furthermore, this study ensures the anonymity, privacy for participants, as Saunders et al. (2009) suggested. If I need permission for anything, I would ask for interviewees' responses.

I try to assure that respondents feel no harm by giving answers to my interview (Eriksson & Kovalainen, 2008). I do this by actively discussing the interview with participants whenever they have any concern and make sure that there is always room for questions and answer. Also, I pay more attention to my communication to avoid unnecessary stress for respondents. Throughout the interaction, for example, when talking about any concern when using mystery shopping results, I try to observe any discomfort or annoyance from the interviewees to make suitable adjustments.

3.8 Evaluation of qualitative research

According to Denzin & Lincoln (2005), Farquhar (2012), the validation of qualitative interpretive research can be assessed by four criteria, including credibility, dependability, transferability, and confirmability.

Credibility: According to Lincoln & Guba (1985), credibility refers to absolute "truth" in my study. This study would ensure a specific agreement between the interpretation

and subject opinions by digitally recording the interviews and engaged in discussing the interviews with respondents.

Dependability: Dependability focuses on the carefulness and consistency of the research (Lincoln & Guba, 1985). This study paper would show the consistency of all informed choices.

Transferability: According to Lincoln & Guba (1985), transferability refers to what extent the study findings can be transferred to a different context. The interviews yield a number that is hard to present for any sector. However, companies still can use the findings as a reference to better understand and improve mystery shopping in their companies.

Confirmability: Confirmability weights how well the findings and interpretations are logical and non-prejudiced linked (Licoln & Guba, 1985). Regarding confirmability, the structure and design would provide clearly references and thus, clarify the respondents' statements, opinions, and interpretations.

4 FINDINGS AND DISCUSSION

4.1 Process to produce mystery shopping results

To facilitate understanding of the context for mystery shopping measurement and how store managers use the results, this part will present the process to produce mystery shopping results at Guardian - the workplace of respondents.

In general, store managers in this study can briefly describe the mystery shopping process at their workplace. Based on respondents' answers, the following figure was built to illustrate the context for mystery shopping measurement with a focus on the store's view.
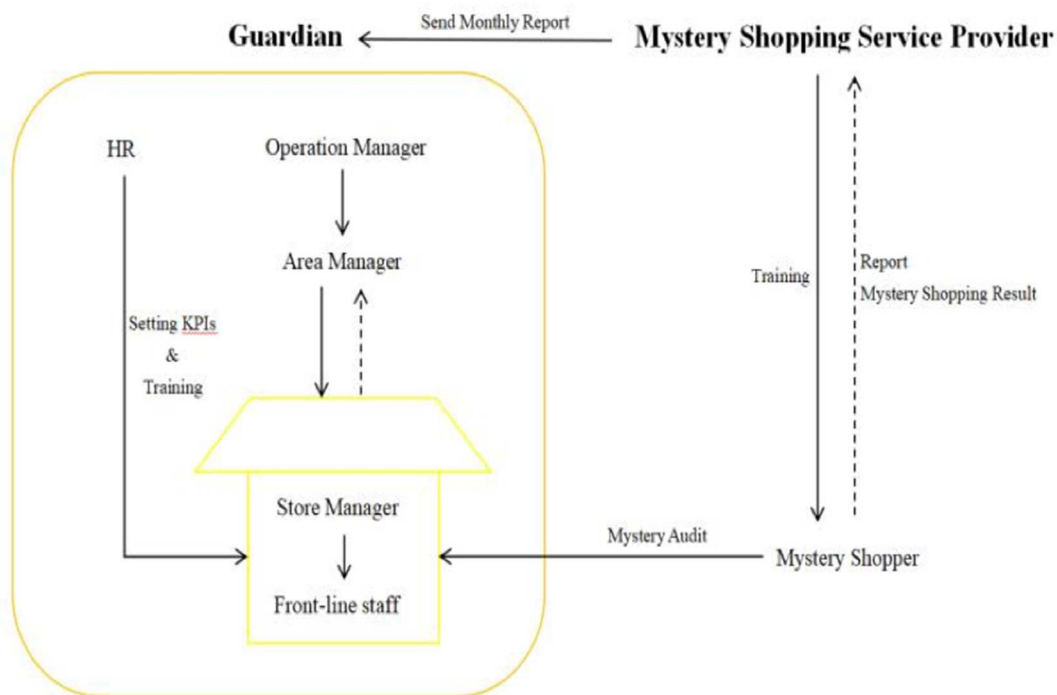
Figure 4: The context for mystery shopping measurement at Guardian

Defining benchmarks

Store managers are not involved in this phase; however, respondents in this study know that their company and mystery shopping service provider would agree upon evaluation and scoring criteria for mystery shopping. Upon agreement, mystery shoppers can evaluate the store performance against pre-determined standards, and provide some comments and scores that reflect their assessment.

> "When my company signed a contract with that company, they had met…so that company must know the dimension of service quality in my company, how to grade." (Respondent E)

Training

<u>At Guardian</u>

According to one respondent who had worked at Guardian long before the company launched mystery shopping, the HR department of Guardian organized training sessions before the mystery shopping took effectively. The training step is crucial to make sure everyone is on board. Frontline staff who serve customers every day are the ones that will be observed and evaluated by mystery shoppers. Hence, it is crucial that the store staff are aware of the program and willing to act upon the results.

Store staff who joined the training would receive information on "why company adopts mystery shopping, what it measures, the scoring system for mystery shopping" (Respondent B). They also received "three or four A4 pages of the checklist for mystery shopping" as training materials (Respondent B).

Another respondent reflected on the training she received as below:

> "When I started working as an Assistant to Store Manager, I learned about the program. The company had a session to talk about it, why mystery shopping, how it helps the business, why the company has to spend a large amount of money for it. Otherwise, employees will misunderstand that the company use it to spy on employees." (Respondent D)

Her response offered insights into some common ideas among staff, such as "company spy on employees," "company catching something wrong." According to this respondent, effective communication about the program purpose is critical to avoid those misunderstanding among staff. Thus, employee training is a useful tool to prevent and reduce misunderstanding about the program.

Despite the importance of training, the rest respondents claimed they did not receive any formal training about the mystery shopping. They learned about mystery shopping from their "seniors," "store manager," "area managers," and their "own experience." This can be explained that due to shift work and tight headcount control at the store, not all staff

can join the training session. Hence, only a few staff were selected to participate in the training, and the participants are expected to spread out the information gained from the training. Besides, while even some of the store managers had no training from the company, I found that respondents generally agree that store managers are expected to train and inform new staff about mystery shopping.

At the mystery shopping service provider

From their own experience as lacking formal training from the company, two store managers expressed their concern whether the mystery shoppers get any proper training from their company to be qualified to evaluate the store. As one respondent concerned:

> "I know briefly then I tell my staff … based on the checklist, the criteria in that checklist. So I doubt if mystery shoppers were trained or not. Or they also get a checklist and go on evaluation without understanding it." (Respondent A)

Most respondents still agree that mystery shoppers get trained to know about the criteria and scoring system. However, some respondents doubt the qualification of mystery shoppers to a certain extent, and more information about this concern will be presented in later sections.

Mystery shoppers visit the store and report the results

Mystery shoppers visit each store in the retail chain two or three times a month. They evaluate the store based on the pre-determined checklist. By the end of the month, mystery shopping service provider would deliver a mystery shopping report to Guardian and store managers would receive the report from their direct managers, area managers. The following section will outline the key components of the mystery shopping report.

4.2 Components of mystery shopping report

The mystery shopping report includes two main parts: general evaluation and stores' detail evaluation. General evaluation summarizes the evaluation of the performance of

all stores in the retail chain. This part of the report also shows the score of each store. There are assigned points earned for each behavior measured.

> "For example, when customers enter the store, we must give the basket, greet them and actively ask what products customers are looking for. If one of the three steps is missed, the points will be deducted." (Respondent F)

The total points earned will be divided by the total point 100, yielding a percentage of points earned relative to the scale of 100. A score above 80% is considered outstanding performance, score from 76% to 80% is regarded as good performance and below 76% indicates the store need improvement.

In the second part of the report, stores' detail evaluation provides evaluations on environment and facility, and frontline staff. Environment and facility evaluation focus on in-front and in-store evaluation. While staff evaluation focuses on staff grooming, staff attitude, and manner, the staff at the cashier and end of the transaction, other staff at the store, staff with outstanding performance.

This study aims to find out what store managers do after receiving the mystery shopping report. Hence, the following part will present the findings from the interviews with store managers, focusing on what managers do with mystery shopping results and factors influencing their behaviors. I will first give an overview of how store managers approach the mystery shopping results and then discuss store managers' behaviors and influencing factors.

4.3. The overall approach to mystery shopping results

In this study, most respondents claim they use mystery shopping results as a reference to know what the store has done well and what needs to be improved. However, all respondents spend more time to talk about how mystery shoppers help identify something wrong at the store rather than what done well. This is in line with findings of

Wilson (2001) that mystery shopping can help the organization to identify weaknesses and failures in the frontline service.

> "It is helpful in 2 ways. First, it helps further develop the strengths of that store. Second, it finds out weaknesses of that store to help the store better. Sometimes, even you are available at the store; you still cannot see how bad the staff attitude or services are." (Respondent F)

The response above shows that store manager acknowledges the usefulness of having mystery shopping like a third party to check the service quality at the store. Despite being often present at the store, being too busy or too related to the context can blind store managers to have an objective view of the store's service. Store managers may not be aware of "how bad the staff attitude or services are." So thus, mystery shopping is needed to look at the problem from outside. Another respondent makes reference to using mystery shopping results to prevent adverse outcomes such as "customer complaints" or "customers choose another store."

> "Maybe customers are not satisfied with the service, but they will not tell us. Mystery shopping helps us to identify the problem before it gets worse that customers may leave us." (Respondent D)

The following quote, from another manager, provides more details on how mystery shopping helps the manager check and become aware of some bad performances that she was not aware of in her store:

> "There are small issues that I have never thought about, but mystery shopper found out…unexpected things … Simple things, for example, the gap between cosmetic selves…a small gap, there is dust inside. No one knows…everyone is busy…cleaning the products takes much time. However, mystery shopper found out and took the photo. Have to say they are good." (Respondent C)

The response above refers to failures identified by mystery shoppers as "simple," "small." This implies that those mistakes are not important to that store manager. As the store manager explained, staff were busy with cleaning the products and thus, have no time to pay attention to that detail. However, she is still glad that mystery shopper has helped find that "unexpected thing" and thus, her store can improve the service and maximize the performance. In the interviews with other respondents, I also found that some acknowledge the use of mystery shopping in identifying "minor," "simple" mistakes.

As the frontline staff' must follow the framework regulations, there is a relatively little reference that mystery shopping results are used as a means of controlling standards. As store managers bear responsibility for keeping daily operations running smoothly, one respondent refers to the mystery shopping report as "a daily store checklist that checks and ensures the store function well." (Respondent G)

Overall, the findings provide the idea that store managers adopt mystery shopping results for a variety of purposes ranging from controlling standards to maximize performance. Despite acknowledging the mentioned use of mystery shopping results, all respondents are still much concerned about the reliability of the results. As a result, store managers "could not use the results to solve store problems or improve store performance if it was not true" (Respondent F). The interview transcriptions offer references that respondents rely on their intuition and work experiences to analyze and interpret the results. Some of the respondents make detailed references to analyzing and taking further action base on the information from the mystery shopping report. The following sections will provide more significant details in each step of store managers.

4.4 Analyzing mystery shopping results and influencing factors

This section will provide more details into how store managers analyze the mystery shopping results. It is found that store managers categorize the results to those they found "reliable," "reasonable" and those not. Two respondents share their ideas on what

information they find useful and how they analyze the results as below. In their opinions, some criteria cannot be "evaluated correctly" by mystery shoppers. Hence, it is suggested that store managers only adopt the results on criteria that they found reasonable and keep in mind that the results only reflect a partial view of the store's operations.

> "I found staff attitude, and image display information are useful. I think if I step into the store and stand at the selves, if the selves are too dirty, that's too bad. So we can get the feedback on those and get better. However, based on that and judging the store managers, the employee…I think it only reflects about 50 % of the situation." (Respondent F)

> "Mystery shoppers reflect the positive and negative sides of the store. However, it depends on the criteria they are evaluating. For example, employee attitude, can a secret customer be able to assess it? For example, the report says the staff unfriendly. However, how unfriendly is unfriendly?" (Respondent D)

The quotes above reflect the issue of what can get measured by mystery shoppers. The problem here is that the service sector consists of tangible and non-tangible elements, and has the so-called HIPI characteristics. That service is heterogeneity, intangible, perishability, and inseparability and thus, measuring the service quality is a great challenge for both company and mystery shoppers. In general, criteria on the store physical and products are more natural to be measured and evaluated while the human element is harder to be assessed on a fair basis by mystery shoppers. In this case, employee attitude is not a tangible and measurable characteristic of employees. As one respondent states, "it only reflects about 50% of the situation." Hence, there is an observation that all respondents became careful when analyzing the results that involve human factors since most managers do not trust those results.

Many references show that most store managers do not trust the results as valuable, reliable to a certain extent. More specifically, most respondents find the comments on

non-tangible elements are "general statements" and lack of evidence as one respondent claims.

> "I do not believe those results, because firstly I do not have a specific score, the second is there's no image… the result I receive, no evidence to prove that part." (Respondent A)

The reliability of the mystery shopping results is often linked with the reliability of mystery shoppers who gather the data. There are some respondents expressed their concern about the quality of mystery shoppers in terms of observing and providing a fair evaluation.

> "It's about their profession, and if they want to survive in their business, then they have to point out the weakness of the service. They can't always evaluate a store as good… Maybe their company also sets goals for them to achieve, so what if they create fake evidence?" (Respondent A)

> "How they train mystery shoppers?…maybe those are just students do part-time. The quality of the mystery shopping…good or not? Maybe my company requires the quality, but we can't verify it…we just read the report every month." (Respondent C)

The first response above reflects a common perception among staff is that mystery shopper "catch them doing something wrong" rather than "catch them doing something right." This stigma even leads the respondent to doubt if mystery shoppers create fake evidence to fulfill their assignments. According to respondent D, since "store managers have no way to check how mystery shoppers produce the results," store managers then base on the report to evaluate the qualification of mystery shoppers. Wrong information in the report further reduces the store managers' trust in mystery shoppers. In interviews, most respondents can provide several references that mystery shoppers provide wrong evidence.

"For example, mystery shopper identified an employee in my store with that mistake; however, the employee's name is wrong. It's the name of the staff in another store, not my store. However, my store's score was deducted." (Respondent F)

There is still one respondent who have a more positive view about the reliability of mystery shopping method. Respondent B reflects on her experience with mystery shopping and sees mystery shopping has improved as "the evaluation criteria has changed … they removed some unnecessary ones; in general, it is easier to manage. " The improvement makes her attitude shift to be more positive towards the program. "I feel it's positive. Lately, I feel it has changed, so I feel better. It was so bad before…Now it's more objective, and the results include more photos."

Dealing with those "vague, " "unreliable" results, most respondents have to rely on their intuition, and work experiences to decide which results are reliable, which mistakes identified by mystery shoppers are real problems that need to work on. Also, when the mistakes identified by mystery shopper seems right and reasonable to store managers, most respondents would find the cause of the problem. Several of the respondents reflect on what they perceive as how mystery shopping results only convey a limited view of store operations.

Most respondents often refer to those factors that "beyond their control" or "unreasonable, " "unlucky" as justification for some mistakes identified by mystery shoppers. They tend to understand why the staff made those mistakes, and thus, can accept those mistakes as cannot fixable.

"Many times, when it is too crowded, it is impossible to serve customers, there are days we have 300 customers." (Respondent F)

Organizations aim to maximize both sales and customer service since excellent customer service is believed to lead to increasing sales. However, in this study, most respondents acknowledge that there is a trade-off between sales and customer service to a certain extent, especially when the stores have to serve too many customers. As

Guardian specializes in health and beauty products, the staff are much more expected to provide consultation and advice on products and beauty tips. Compare to other retailers that sell products and do not need much consultation such as groceries, the pressure for Guardian staff is pretty much.

According to Wagnerová & Baarová (2008), the pre-determined criteria in the checklist indicate preferred behaviors among employees by the management board. One respondent reflects on many desired behaviors by the company. In her opinion, it is hard for the frontline staff to follow all the procedures, especially when the store is crowded and busy.

> "Price tag, service attitude, greeting customers…too many regulations." (Respondent D)

Another respondent provides further details on some "unlucky and encounter mystery shopper" mistakes. When the store is in some situations that staff cannot follow all the procedures under time constraints, such as being crowded, her staff have no option rather than "do right but not follow that details." It is hard for store staff to keep up with expectations from the company to follow all the procedures and regulations. In case the staff is spotted by a mystery shopper, that would be considered as "unlucky," according to that respondent.

> "If we are unlucky and encounter the mystery shopper, they learn that we have to introduce five programs, but we introduce four; then scores are deducted. Sometimes the store is so crowded, some staff remembers, some not. At the counter, dealing with a long queue, staff have to give change with two hands, say 'Thank you and see you again,' if we just say 'thank you' we will be penalized for not saying 'see you again'." (Respondent E)

Store managers are often visible on sales-floor to make sure things go smoothly. Hence, they are often get updated about the status of products, staff, and the overall store. Gradually, they became aware of some factors that may affect the mystery shopping results. In retailing service operations, besides some simple tasks that easy to get

managed, there are many intangible and unforeseen issues that outsiders like mystery shoppers cannot understand. Another respondent reflects on this issue as below.

> "When customers enter the store, we need to invite them to use their cell phones to scan the QR codes to get promotion information and the leaflet. We will be penalized if not do it. I know now we have to follow technology… however, many customers do not like it. Those housewives…in forties…it is inconvenient for them, sometimes we introduce that and other stuff…many stuff, then customers get annoyed and ask us, 'What…so what do you want?'." (Respondent D)

The response above shows the store manager's frustration when her staff has to balance between following the procedures and personalizing the service to please customers. It can be said that the work is heavily prescribed for store staff and one-size-fits-all approach from the company sometimes put the staff in a dilemma. The mystery shopping results, thus, are not welcomed by most respondents. Because it cannot "empathize" with challenges and difficulties that staff encountered at the store.

One respondent summarizes the preferred way to identify and analyzes problems then point out the weakness of the mystery shopping method is looking at problems too narrowly with a "single view." Store managers, with their sharp insight in the filed, are aware of some factors that may affect the accuracy of the report. They consider those factors when analyzing the results.

> "I think that when you look at a problem, you have to look from many perspectives. It is possible that the store is changing its layout, or maybe there is a problem, or there are too many customers, too busy, the store lacks employees. There are many reasons. That mystery shopper only has a single view, instead of the multi-dimensional view, so the things he evaluates, even if captured with images, it only reflects 50 to 70 percent of the problem." (Respondent E)

4.5 Interpretation of the results and influencing factors

According to Bourne et al., (2005), interpretation is drawing out meaning from the performance measurement system. In terms of mystery shopping results, interpretation of the results is made by checking the mystery shopping score and comparing the score with the benchmark. In this study, all respondents check the mystery shopping score and base on that to take further actions.

When the score is below 76%, all respondents would take it seriously and focus on improving store performance. Store managers know that below 76% means their stores are at the bottom of the ranking and need improvements. Getting a low score also means that their area managers would pay more attention to their stores. Most respondents state they need to concern about low score since area managers would email the store, requiring explanation and taking action. Then store managers have to give explanation and promise for improvements next month. This can be illustrated with the following quote:

> "Low score is…tiring, we need to list out mistakes, commit that the store will improve somehow. The score just decreases 1, 2 points [compare to last months] is enough to get trouble with area manager." (Respondent B)

Besides the pressure from the area manager, stores also have to deal with the HR department when they have low scores. Specifically, when the score continuously below 76% three times, the store will receive a warning letter from the HR, and it will affect store managers' KPI evaluation in the year-end performance review. For those stores that often rank at the bottom, sales staff of those stores will be appointed by HR for further training classes. Hence, most respondents would try to keep the mystery shopping score of at least 76%. When were asked how about aiming higher goal, most respondents show a lack of motivation and enthusiasm for that. As one respondent explains the score highly depends on mystery shoppers.

"That's difficult…because it involves human. Who knows when mystery shopper is happy or not. If they are happy they can give a good score, if they not, they give bad score." (Respondent D)

The mystery shopping results are influenced by the behavior and interaction of both participants – mystery shopper and the staff. However, in the view of respondent C, the results depend more on mystery shoppers than on staff, as she further explains. By analyzing the pattern of scores across the chain, this respondent forms the belief that the score depends on characteristics (easy-going or strict) and the mood (happy or moody) of the mystery shopper.

"I check the score of all stores and realize those stores close to each other will have similar scores. Whenever my store gets a low score, those stores nearby also get low scores. That because mystery shopper will go and check stores in the same area. When she is happy or easy-going, all store will have high scores." (Respondent D)

Since the report shows the scores of all stores, store managers can check and compare their scores with other stores within a chain. As mentioned before, there is a perception among store managers that there is a trade-off between sales and customer service. Thus, according to another respondent, those stores with low sales volume will have high mystery shopping score and vice versa. In the view of store managers, there is a lack of understanding or empathy by the company and "higher managers," of the store constraints and daily challenges encountered by the frontline.

"Those stores with low sales volume will have high mystery shopping scores, those store in remote areas…also high scores. As for the crowded stores, it's scarce to see they have a high score. The reason is simple … too crowded. Those empty stores, nothing to do, they are too happy to meet customers [laughs]. When there are not many customers, if a customer needs, I also can spend a half-hour to consult that customer." (Respondent E)

Guardian uses mystery shopping results to benchmark stores against each other. The principle behind this is that all stores in the retail chain must provide equal service and products to maintain a consistently strong brand identity. However, for some respondents in the study, this measurement was perceived as unfair since it failed to capture the unique constraints in each store environment.

There is a recognization among respondents that the employee reward and recognition system linked with mystery shopping seems not effective and efficient to motivate staff. Guardian has a reward for the store that maintains three months continuously has the highest score. However, most respondents see that goal as impossible. Since all the managers I interviewed have not seen any store makes it. "High score… then you can get something, but it's rare." (Respondent E). Most respondents show no interest in getting the reward that they see unreachable, and some even do not know precisely what is the reward.

At the store level, to motivate staff, one respondent share how she treats the staff food and drink whenever her store has a high score. In this way, getting a high score makes her cost money, and as an exchange, she can avoid listening to the manager's complaint.

> "Any month we receive a high score, we will eat and drink, reward our store. Every month that my store scores over 80… there's eating … costly…waste of money. I see being store manager is costly…However, with a high score, you don't have to listen to complaints from the area manager." (Respondent B)

Some respondents also mention the issue of identifying mystery shoppers in their interviews. The logic of mystery shopping is that the identity of a mystery shopper is not revealed to staff and organizations should make spotting mystery shoppers less important by promoting the usefulness of this method. However, from the view of one respondent, spotting mystery shopper can affect the objectivity of the report. This perception also implies that mystery shopping result is not reliable for store managers, no matter the results are a good or bad, low, or high score.

"Honestly, no one would support mystery shopping…simply because…in store managers' opinions, it's meaningless. For example, if I recognize mystery shopper, I behave to be very good and considerate…Have to say it's a bit fake, but after mystery shopper leave, is it guaranteed that the store will treat other customers like that?" (Respondent C)

In general, in some store managers' mindsets, getting a high score does not necessarily mean that store performance is good. Most respondents show no interest in aiming high as they see it is unreachable to get a reward liked with mystery shopping results. Most respondents treat the high score as a way to avoid the problem with their direct manager-area manager. In a way, this suggests a form of dysfunctional behavior, convergence, as store managers tend to stay within recognized limits and have no motivation to be outstanding. Besides, the implication that staff can identify mystery shoppers and purposely treat the mystery shoppers well can be understood as the behavior of gaming. That sales staff can manipulate their behavior to generate the outcome.

Compare to when the results are good, store managers have a lot more to do when the results are not good. A manager, in her interview, describes what she does after receiving the report, and the results are not good.

"Make a list of our store's mistakes then disseminate with my staff. If other stores make some serious mistakes that area manager mentioned, I also list those mistakes. If the store not good, we need to improve." (Respondent F)

After receiving the mystery shopping report, all respondents then would briefly review the results with their staff. This communication is vital since the sales staff are those being evaluated by mystery shoppers, and they need to "know what they did wrong so that they can fix" (Respondent F).

4.6 Dealing with staff and influencing factors

Since the mystery shopping report provides much information about the staff. All respondents show their recognition of human factors and carefulness when assessing human flaws. Some respondents express how they put themselves to the staff position to see the mystery shopping criteria is reasonable or not. This suggests the balance between the role of a store manager and as a human being. The following quotes express managers' frustration with irrational criteria in the mystery shopping checklist.

> "Mystery shoppers also evaluate the guard service in front of the store. However, I don't see it as good. The score for that part accounts for a very high percentage… I think that the guard just needs to be friendly, keeps the motorbikes safe, assists customers in parking…But the working time is very long, from 8 am to 10 pm, most guards smoke a lot. Now, smoking is not accepted; using the phone is also photographed [by mystery shoppers]. …God, if I were the guard…I can't stand it… It is not reasonable. So are we too strict with him or not?" (Respondent G)

> "Mystery shoppers go to the store, ask questions… they look for …too much detailed answer. For example, some ask the staff how to use the product … that's ok, the staff knows, staff can tell how to use, the origin of products … these are normal. But you ask the staff more about detailed ingredients…who knows…too intensive." (Respondent C)

The first response reflects how the store manager feels challenging to balance her role as the store manager and as a normal person. Should she be strict with the wrong behaviors of the security guard? Or she accepts those, but those mistakes will affect her store performance and her KPI on mystery shopping. Weighing between two options is not easy for store managers. Most respondents in this study seem more flexible to those mistakes as they see some evaluation criteria are unreasonable as discussed before.

In case the store is underperformance and get a low score, some respondents claimed they find it difficult to "talk seriously" or discipline the staff if base solely on mystery shopping results.

> I ask my staff, and the staff says 'I did well, I don't know,' but mystery shopper's evaluation is different … how can I know who's right, so I can only tell staff to improve." (Respondent F)

This response shows the difficulty that the store manager encounters when communicating with their staff about the bad results. Unreliable data makes this respondent not confident to take further actions. Even with further investigation, this store manager still cannot decide which side is right, and thus, she choose to compromise by merely encouraging the staff to behave better next.

Another manager reflects on her experience in dealing with her staff about the bad results. This respondent finds the problem lies in the staff attitude. In her opinion, if the staff has a good attitude, store manager can tell them and trust that they can improve. However, for those with bad attitudes, store manager cannot trust those staff.

> "Regarding the attitude of the staff, there is some nice staff, and then they will improve…However, some stubborn…they nod their head but will not act accordingly …What can I do … I can't follow the staff 24/24." (Respondent E)

When the store is underperforming, most store managers state they can feedback and discuss the results only with staff who have good attitudes and manners. Because according to some respondents, even though the training may help improve the qualification of the staff, it is hard to improve the staff attitude. Those staff who commit to work, want to be promoted, their attitude is more positive than negative, and thus, their behaviors at work would be better than those with bad attitudes. Store managers can feedback active staff with the mystery shopping results, and thus, they can help improve the store performance.

"There are two types of employees, first are those work temporarily, they just work so so … don't need to try hard. If the staff want to work harder, get promotion, then they are the second type, their attitude will be different, they think differently." (Respondent F)

Store managers have difficulty in dealing with their staff not only because of human factors but also because there is no clear evidence for the staff's mistakes. A number of store managers in this study complaint about lacking evidence when mystery shoppers feedback on tangible issues such as staff attitude, counseling skills.

"Why the store have been deducted from counseling points, there is no picture, explanation … the same for service, attitudes of employees." (Respondent G)

"For example, mystery shopper commented that the staff provided wrong counselors. How wrong is wrong … They just say and say general statements." (Respondent D)

The meaning behind performance measurement is that managers can use the results to take further actions and drive improvements. Store managers are responsible for managing staff and providing appropriate training, coaching, or discipline to ensure the well-being of the store and sales staff. However, the relationship between the store manager and staff is fragile. In view of some of the respondents, there is a danger of encouraging unintended behaviors by putting too much pressure based on underperformance. Few examples are presented as below.

"Some staff, they have the idea that they don't like this place, they can work elsewhere...all of them are the same. There are new places with no pressure, less pressure…then they can move there." (Respondent F)

"The staff take care of customers more than me. So if they become so stressful. It's also not good. If their mood is not good, how can they serve customers." (Respondent C)

Most frontline staffs in retailing are low-paid. Hence, store staff is easy to switch to another company that offers a higher salary or a better job. According to respondent G, "incorrect mystery shopping results may lead to poor staff satisfaction." Together with pressure from the company about the underperformance, incorrect mystery shopping results may increase the rate of turnover. Besides, pressure from mystery shopping program may influence the staff mood. Among respondents in this study, it is well recognized that sales staff are important as those contact with customers on a daily basis. Without sales staff, the store cannot function at all. Hence, some store managers show their sensitiveness about the well-being of sales staff, and thus, they try to avoid unnecessary stress for staff from the mystery shopping results.

4.7 Dealing with Area Managers and influencing factors

> "Given mystery shopping result every month, the area manager will raise the issue of a low score, then stores know what needs to be done. Higher managers don't need to care about the results since area managers care enough…Even no need area managers, store manager must know what to do." (Respondent B)

As mentioned before, store managers would try to get a high score as a way to avoid troubles with their area manager. In the quote above, the respondent shows that most store manager even stays alerted ahead of their area managers. Most store mangers know that their area managers are not pleased with bad results. Some respondents reflect on the way the company top-down the mystery shopping program and do not provide enough support for store managers. As a result, each store manager has her own way to review performance, determine areas of improvement, and come up with further actions.

> "They just top-down, saying that it's good for the development of the store, they want the best service, based on these criteria. How to ensure these criteria are not known. People work on it by themselves. Each store manager has their own way." (Respondent A)

Given the trade-off between sales and customer service to a certain extent, there is a perception among most respondents that area store manager would be easier on the store if the store still guarantees good sales and achieves financial KPIs. As one respondent states, "as long as the sales volume is good, most mistakes in the mystery shopping report can be forgiven" (Respondent B). This implies the less importance of mystery shopping KPI, compare to financial KPI and thus, mystery shopping KPI seems less effective in influencing store manager's behaviors. Some store managers admit they are less pressured with mystery shopping results when the store keeps good sales performance.

Despite much reflection on the unreliable results, there is an observation that most respondents accept the results. Most respondents would not feedback officially to their area managers about their concerns regarding mystery shopping results, but rather complaint with area manager in some ways. One reason is the lack of support from their direct managers. All respondents state that their area manager would suggest them to look on positive sides and focus on improving the store performance. This can be illustrated with the following quote:

> "In general, there are many annoyances about the program, but instead of expressing, I should look at them and make improvement. That is what the bosses think." (Respondent B)

Another manager reflects on what her manager told her.

> "According to my manager, I should not take the results too serious. I should see mystery shopper as a usual customer. So if I follow the company regulations, I don't need to be afraid of mystery shoppers." (Respondent E)

Although acknowledging the advice of area manager is true, respondent E still feels challenging to deal with inaccurate results. This store manager has been promoted from sales staff to assistant store manager then to store manager. Thus, she understands that a higher position means more responsibilities and need to look at mystery shopping

results in a way as her area manager suggests. She reflects on how her ideas about mystery shopping has changed gradually as below.

> "When you are promoted, you have little responsibility…Then you must try harder, be more responsible, and you understand better. While staff' mindset is simplier." (Respondent E)

Only one respondent in this study used to feedback about the mystery shopping program with her area manager. However, she did not receive any response, and since then, she feels feedback about mystery shopping results is "just a waste of time."

> "Once, the staff took a picture of the mystery shopper to feedback, because it had many unreasonable, unreliable points. But the point of my feedback… is that are they using low-qualified people to evaluate the store? This mystery shopper is too bad … she plays the role of a secret customer but then discovered by staff … then was photographed. I send it to the company, but I got no reply." (Respondent A)

Another respondent explains the acceptance of mystery shopping results as "we can't change it, so we accept it" (Respondent D). This response reflects the passive state of store managers when they feel mystery shopping program is top-down, and they have limited control over the situation.

While most respondents regard the feedback as useless, and they feel the company does not welcome their voice. Another respondent explains why she would not feedback about the results even though she feels the results is 'unreasonable' as below.

> "Feedback is rarely recorded, feedback is troublesome. I think it is also not worth it. For example, depending on the criteria and the score…For instance, my store's score is 65, if I feedback and I can increase the score, but my score is still below 76. That's useless." (Respondent C)

There is an implication here that the scoring system discourages the feedback. Store managers are busy with many responsibilities, and thus, they would not waste time for

feedback if it is not worth. Another reason for the accepting attitude towards mystery shopping results is the service quality orientation at Guardian. All respondents acknowledge the need for mystery shopping since they are working in service filed. Even though some mystery shopping seems unreasonable, most store managers still accept it as they would consider that "customers always right." There are a variety of references about the importance of service quality, customer service as perceived by store managers.

> "We sell products that others also sell, we only better at customer service. Customer service is what we have…we serve people, so we have to listen to customer's opinions to improve our services. " (Respondent G)

> "In general, the customer is always right, so when I get the information, I first tell the staff to improve… Generally, that's service." (Respondent C)

Another respondent shares her evaluation of the service quality at Guardian, base on the comparison with other retailers' service. The store manager can feel the importance of service quality at work and thus, better understand the need for mystery shopping results to improve the service quality and customer service.

> "That the service here is so important, the service is so important that when I go to other retailer stores, I see their service is bad…so I can compare; there's a bit called I proud of the service here." (Respondent F)

5 CONCLUSIONS

Since assessing performance and service quality in retailing has become more important than ever, retailers are keen on using mystery shopping as a method to evaluate customer service and frontline staff. Organizations use mystery shopping reports for a variety of purposes, mostly to identify good and bad points in their service, to compare the effectiveness of each store as well as to ensure the standards across the retailer chain. However, there is limited insight into what happened after the managers receive the mystery shopping reports

This study aims to explore what store managers do with mystery shopping results and any factors that influence their behaviors. The study uses a qualitative approach, and data was collected by semi-interviewing with seven store managers from the same retailer. Qualitative data from interviews were analyzed answer the following specific research questions:

1. What do store managers do with mystery shopping results?

2. What are factors influencing store managers' behaviors toward mystery shopping results?

The collected data was positioned within the process stages for performance management by Bourne et al. (2005), including data analysis, interpretation, communication, and taking actions. The findings provide the idea that store managers use mystery shopping results for a variety of purposes ranging from controlling standards to maximize performance. Despite acknowledging the mentioned use of mystery shopping results, all respondents are still much concerned about the reliability of the results. As a result, store managers cannot use unreliable results to solve store problems or improve store performance.

The interview transcriptions offer references that store managers rely on their intuition, perceptions and work experiences to analyze and interpret the results. In their opinions, some criteria cannot be "evaluated correctly" by mystery shoppers. Criteria on the store physical and products are easier to be measured and evaluated while the human element is harder to be assessed on a fair basis by mystery shoppers. Hence, it is implied that store managers only adopt the results on criteria that they found reasonable and ignore those "beyond their control" or "unreasonable," "unlucky." Store managers also keep in mind that the results only reflect a partial view of the store's operations.

In this study, interpretation of the results is made by checking the mystery shopping score and comparing the score with the benchmark. When the score is below 76%, all respondents will take it seriously as they have to deal with pressure from their direct managers – area managers and HR to focus on improving store performance. It is found

that, in some store managers' mindsets, getting a high score does not necessarily mean that store performance is excellent. Since there is an implication that staff can identify mystery shoppers and purposely treat the mystery shoppers well. This behavior can be considered as a form of dysfunctional behavior, gaming.

Furthermore, most of the respondents perceive getting a high score as a way to avoid the problem with their area managers. However, store managers show no interest in getting the reward liked with mystery shopping as they see it unreachable, and some even do not know precisely about the reward. This implies that store managers tend to stay within acceptable score range than be outstanding with the highest score.

Most information on the report involves the frontline staff, and thus, dealing with employees is one of the critical action of store managers after receiving the report. The study found that store managers all communicate the results and most find it challenging to have a serious talk or discipline staff when the store is underperformance. Store managers have difficulty in dealing with their team not only because of human factors involved but also because there is no clear evidence for the staff's mistakes. Together with pressure from the company about the underperformance, incorrect mystery shopping results may increase the rate of turnover. Hence, some store managers show their sensitiveness about the well-being of sales staff, and thus, they try to avoid unnecessary stress for staff from the mystery shopping results.

Although all respondent complaints about the results to a certain extent, most store managers accept mystery shopping results and rarely feedback about the program to higher managers. This can be explained as store managers feel they lack support from the company, and they all aware of the service quality orientation in their organization. As store managers get progress in their careers and bear more responsibilities, they can understand the need for mystery shopping for serving customers better and remaining competitive in the market.

5.1 Theoretical contribution

This study seeks to explore what exactly store managers do with mystery shopping results, and thus, contributes to the currently limited literature on the management of mystery shopping results and addresses the call from Radnor & Barnes (2007) for further research from the operational management perspective towards performance management.

Organizations adopt mystery shopping with the hope that performance measurement will encourage positive changes in managers and employees' behaviors, and subsequently, result in customers' satisfaction and better sales performance. However, in reality, several factors may interfere with the link between mystery shopping results and managers' behaviors, and thus, influence the outcome. Several researchers such as Neely et al., 1995; Waal, 2003; Franco-Santos & Bourne, 2005 and Bell, 2010 imply the link between performance indicators and managers' behavior relies on managers' perceptions and interpretations. Also, it is recognized the potential of various factors that may influence managers' behaviors. Similar findings emerge as this study found various factors in the context of store operations that influence store managers' behaviors such as store managers' perceptions toward the reliability, fairness of mystery shopping results, the reward and recognition system, the support from the company, and service quality orientation. Interestingly, the study found a perception among store managers that there is a trade-off between sales performance and customer service to a certain extent. Together with another perception that implies the less importance of mystery shopping KPI, compare to financial KPIs, mystery shopping seems less effective in influencing store manager's behaviors.

Despite acknowledging the role of mystery shopping, most store managers find it challenging to integrate the results with other resources to make positive outcomes. The findings from this study can be seen as more insight to explain why mystery shopping can provide useful managerial implications and actionable recommendations for assessing service performance, coaching employees, and improving sales performance

(Cramp, 1994), but the relationship between mystery shopping results and customers' satisfaction and behaviors is not always clear (Wilson, 2001).

5.2 Managerial implications

The findings of this study provide support for acknowledging the importance of training and reward system, not only for store managers but also for frontline sales staff. Frontline staff is the representative of the company. However, they often receive little attention. It is suggested that more training would be useful to help the frontline staff understand more about the need for mystery shopping. While flexible rewards system can help motivate store manager and staff work together towards the goal.

Besides, great care needs to be exercised when dealing with staff about the underperformance to avoid unnecessary stress that may lead to low job satisfaction and a higher rate of turnover.

Most respondents do not welcome mystery shopping results. Because the program reflects a failure of the measurement system to make allowance for factors that are beyond the control of store managers. This suggests that higher managers should be more sensitive to interpret and analyze the results.

Some store managers feel they are in the passive state as mystery shopping is approached from top to down only. There should be room for store managers to feedback about mystery shopping results. It not only helps improve the quality of mystery shopping but also makes store managers feel more listened and supported by the company.

5.3 Limitation

As a qualitative exploratory study, this study aims to explore store managers' perceptions and behaviors towards using mystery shopping results. However, it has to

be recognized that the findings are somewhat limited and can be biased. Regarding the sensitive topic, respondents often struggle to recall and talk about good things.

All the respondents in this study come from the same retailer. This helps to understand the context for exploratory research. However, a handful number of respondents can only represent a small number of store managers, and the findings can be explained by contextual factors that may not present in other settings.

Researcher interprets the findings, and thus, my understanding, knowledge, and perception can affect the findings.

5.4 Further research

This study focuses on the perspective of store managers only. Therefore, more studies from different angles will be beneficial for a more comprehensive understanding of the management of mystery shopping results. For example, the mystery shopping management by higher positions such as area manager, or operation manager can be a topic. Similarly, store staff perspectives can be studied. More research in other retailer settings and perhaps other sectors also needed. Furthermore, all respondents in this study are women. The gender imbalance in this study may suggest a study from the gender perspective.

APPENDICES

Appendix I: Interview guide

Introduction

My name is Tran My Linh. Thanks so much for taking your time for this interview.

I want to ask you some questions about your experiences and ideas about mystery shopping program and how you use the mystery shopping results. This information would provide more insights into the use of mystery shopping results by store managers and the data would be used for my thesis only.

The interview should take about 45 minutes. I would like your permission to tape record this interview. If at any time you wish to discontinue the recording or the interview itself, please feel free to let me know.

Do you have any questions before we start the interview?

Interview questions

1. Please introduce yourself? Can you share more about your experience as a store manager?

2. When and how did you know about the mystery shopping program?

3. Can you describe the mystery shopping process in your company?

Prompts: mystery shopping report.

4. What role does mystery shopping have in your job?

Prompts: How you use the results for what purpose?

5. Can you tell me about the time when the MS results were useful vs. not useful? What the difference?

6. Did u encounter any difficult when use the results? How did you overcome?

7. How do you think the mystery shopping program can be improved?

8. Is there anything you want to add as you think it would be useful for my research? Anything I didn't mention.

Pace the interview

It seems we run out of time, is it ok if we keep it a bit longer?

Closing

Thanks for sharing your valuable insights. I enjoyed the conversation today. Is there anything else you would like to add?

Thank you for your time.

Follow-up

When you say [term or phrase], what does it mean?

Can you elaborate on this?

Can you give me an example?

Why was that important to you?

How did you feel about that?

So you are saying [paraphrase what interviewee said to confirm or clarify]?

Appendix II: Summary content of mystery shopping report at Guardian

I. General evaluation

II. Stores' detail evaluation.

1. Environment and facility evaluation

       1.1. In-front store

       1.2. In-store evaluation

2. Staff evaluation

       2.1. Staff grooming,

       2.2. Staff attitude and manner,

       2.3. Staff at the cashier and end of transaction,

       2.4. Other staff

       2.5. Staff with outstanding performance.

REFERENCES

Allison, P., 2009. *Mystery Shopper Motivations and the Presence of Motivation Crowding*. PhD. University of Central Florida.

Beck, J. and Miao, L., 2003. Mystery shopping in lodging properties as a measurement of service quality. *Journal of Quality Assurance in Hospitality and Tourism*, 4(1-2), pp.1-21.

Beck, J., Lalopa, J. and Hall, J., 2004. Insuring quality service: Training mystery shoppers. *Journal of Human Resources in Hospitality and Tourism*, 2(2), pp.41-56.

Bell, J., 2010. *Managing performance in retail operations: The store manager view*. PhD. University of Surrey.

Berry, L., 1986. Retail businesses are Services businesses. *Journal of Retailing*, 62(1), pp.3-6.

Bititci, U.S., Carrie, A.S. and McDevitt, L., 1997. Integrated Performance Measurement Systems: a development guide. *International Journal of Operations and Production Management*, 17(5), pp.522-534.

Bourne, M., Kennerley, M. and Franco-Santos, M., 2005. Managing through measures: a study of impact on performance. *Journal of Manufacturing Technology Management*, 16(4), pp.373-396.

Brady, M. K. and Cronin, J. J., 2001. Some new thoughts on conceptualizing perceived service quality: A hierarchical approach. *Journal of Marketing*, 65(3), pp.34-49.

Braun, V. and Clarke, V., 2006. Using thematic analysis in psychology, *Qualitative Research in Psychology*, 3(2), pp.77-101.

Broadbridge, A., 2002. Retail managers: their work stressors and coping strategies. *Journal of retailing and consumer services,* 9, pp.173-183.

Buchner, T. W., 2007. Performance management theory: a look from the performer's perspective with implications for HRD. *Human Resource Development International*, 10 (1), pp.59–73.

Burnside, A., 1994. In-store spies snuff out poor service. *Marketing*, 28 April, pp.32-33.

Calvert, P., 2005. It's a Mystery: Mystery Shopping in New Zealand's Public Libraries. *Library Review*, 54(1), pp.24-35.

Chadha, S. and Kapoor, D., 2009. Effect of switching cost, service quality and customer satisfaction on customer loyalty of cellular service providers in Indian market. *The ICFAI University Journal of Marketing Management*, 1, pp.23-37.

Cramp, B., 1994, Industrious espionage. *Marketing*, 8 August , p.17.

Dabholkar, P. A., Thorp, D. I., and Rentz, J. O., 1996. A Measure of Service Quality for Retail Stores: Scale Development and Validation. *Journal of the Academy of Marketing Science*, 24(1), pp.3-16.

Davies, G., 1993. Is retailing what the dictionaries say it is? *Internationai Journal of Retail and Distribution Management*, 21(2), pp.3-7.

Denzin, N. K. and Lincoln, Y. S. eds., 2005. *The Sage handbook of qualitative research*. 3rd ed. Thousand Oaks, CA: Sage Publications Ltd.

Dorman, K. G., 1994. Mystery Shopping Results Can Shape Your Future. *Bank Marketing,* 26, pp.17–21.

Elliott, K. M., Hall, M. C., and Stiles, G. W., 1992. Service quality in the health care industry: How are hospitals evaluated by the general public? *Journal of Hospital Marketing,* 7(1), pp.113-124.

Eriksson, P. and Kovalainen, A., 2008. *Qualitative Methods in Business Research*. London: SAGE Publications Ltd.

Erstad, M., 1998. Mystery shopping programmes and human resource management. *International Journal of Contemporary Hospitality Management*, 10(1), pp.34-38.

ESOMAR, 2005. ESOMAR World research codes & guidelines: Mystery Shopping. [online] Available at: http://www.esomar.org/uploads/pdf/ESOMAR_Codes&Guidelines_MysteryShopping.pdf. [Accessed 1 May 2019].

Farquhar, J. D., 2012. *Case Study Research for Business*. London, UK: SAGE Publications Ltd.

Finn, A., and Kayandé, U., 1999. Unmasking a Phantom: A Psychometric Assessment of Mystery Shopping. *Journal of Retailing*, 75(2), pp.195-217.

Flamholtz, E.G., Das, T.K. and Tsui, A.S., 1985. Toward an Integrative Framework of Organisational Control. *Accounting, Organisations and Society*, 10 (1), pp.35-50.

Forza, C. and Salvador, F., 2000. Assessing some distinctive dimensions of performance feedback information in high performing plants. *International Journal of Operations & Production Management*, 20(3), pp.359-385.

Franco, M. and Bourne, M. C. S., 2003. Factors that play a role in managing through measures. *Management Decision,* 41(8), pp.698 - 710.

Franco-Santos, M. and Bourne, M., 2005. An examination of the literature relating to issues affecting how companies manage through resources. *Production Planning and Control,* 16(2), pp.114-124.

Gagnon, J. L.,Kleinberger, H. and Morrison, G. P., 2005. The customer-centric store: delivering the total experience. *European Retail Digest*, 46 (9), p.76.

Ghosh, D. and Lusch, R. F., 2000. Outcome effect, controllability and performance evaluation of managers: some field evidence from multi-outlet businesses. *Accounting, Organisations and Society,* 25, pp.411-425.

Gronroos, C., 1982. *Strategic Management and marketing in the service sector.* Helsingfors: Swedish school of Economics and Business Administration.

Grugulis, I., Bozkurt, O. and Clegg, J., 2010. 'No place to hide'? The realities of leadership in UK supermarkets. In: I. Grugulis, and O. Bozkurt, eds. 2011. *Retail works.* Hounds-mills:Palgrave Macmillan. pp.193-212.

Guerrien, A., Leconte-Lambert, C. and Leconte, P. ,1993. Time of day effects on attention and memory deficiency: is chronopsychology a method for studying the function of the human subject? *Psycholgica Belica*, 33, pp.159-170.

Hesselink, M., and Van der Wiele, T., 2003. *Mystery Shopping: In-Depth Measurement of Customer Satisfaction*. Rotterdam: Erim Report Series Research in Management.

Kerssens-Van Drongelen, I.C. and Fisscher, O.A.M., 2003. Ethical Dilemmas in Performance Measurement. *Journal of Business Ethics*, 45 (1-2), pp.51-63.

Kompier, MAJ., 2006. The "Hawthorne effect" is a myth, but what keeps the story going? *Scand J Work Environ Health*, 32(5), pp.402-412.

Kotler, P., 1973. Atmospherics as a Marketing tool. *Journal of Retailing*, 49(4), pp.48-64.

Lincoln, Y. S. & Guba, E G., 1985. *Naturalistic Inquiry*. Newbury Park, CA: Sage Publications.

Lowndes, M. and Dawes, J., 2001. Do distinct SERVQUAL dimensions emerge from mystery shopping data? A test of convergent validity. *Canadian Journal of Program Evaluation,* 16 (2), pp. 41-54.

Lynch, R. L. and Cross, K. F., 1991, Measure up - The essential guide to measuring business performance, Mandarin: London.

Marschan-Piekkari, R. and Welch, C., 2004. *Handbook of Qualitative Research Methods for International Business*. Edward Elgar. Cheltenham : Northhampton.

Marschan-Piekkari, R. and Reis, C., 2004. Language and languages in cross-cultural interviewing. In: R. Marschan-Piekkari and C. Welch, ed. *Handbook of Qualitative Research Methods for International Business*. Edward Elgar. pp.224-243.

Martinelli, E., and Balboni, B., 2012. Retail service quality as a key activator of grocery store loyalty. *The Service Industries Journal*, 32(14), pp. 2233–2247.

Mason, G. and Osborne, M., 2008. Business strategies, work organisation and low pay in United Kingdom retailing. In: C. Lloyd, G. Mason and K. Mayhew, eds. *Low-wage work in the United Kingdom*. New York: Russell Sage Foundation, pp.131–167.

Maxwell, J. A., 2005. *Qualitative research design: An interactive approach.* 2nd ed. Thousand Oaks, CA: Sage. Present and future Management Decision.

McCraken, G., 1988. *The long interview.* Sage, Newberry Park, California.

McDougall, G.H.G. and Levesque, T.J., 1994. A revised view of service quality dimensions: an empirical investigation. *Journal of Professional Services Marketing*, 11(1), pp.189-209.

Morgan, N. A., Anderson, E W. and Mittal, V., 2005. Understanding firms' customer satisfaction information usage. *Journal of Marketing,* 69(3), pp.131-151.

Morrall, K., 1994. Mystery Shopping tests service and compliance, *Bank Marketing*. 26 (2), p.13.

Morrison, L. J., Colman, A. M., and Preston, C. C., 1997. Mystery Customer Research: Cognitive Processes Affecting Accuracy. *Journal of the Market Research Society*, 34(2), pp. 349-361.

MSPA, 2018. Mystery Shopping - how big is the market. [online] Available at: https://www.mspa-ea.org/news/newsitem/58-mystery-shopping-how-big-is-the-market.html. [Accessed 1 May 2019].

Myers, W., 2007. *Shopping the shopper: Retail surveillance and performances of Consumerism*. Southern Illinois: University Carbondale.

Neely, A.D., Gregory, M. and Platts, K., 1995. Performance Measurement System Design: A Literature Review and Research Agenda. *International Journal of Operations and Production Management*, 15(4), pp. 80-116.

Neely, A.D., 1998. *Measuring Business Performance: Why, What and How*. London, UK: The Economist and Profile Books Ltd.

Otley, D.T., 1999. Performance management: a framework for management control systems research. *Management Accounting Research*, 10(4), pp.363-382.

PamInCa., 2009. *The Essential Guide to Mystery Shopping.* Silicon Valley: Happy About, p. 119.

Parasuraman, A., Zeithaml, V. A., and Berry, L. L., 1988. Servqual: A Multiple-Item Scale For Measuring Consumer Perceptions of service quality. *Journal of Retailing*, 64(1), pp.12-40.

Parker, L., 1988. Mystery shopping for clues to improve service, *Bottomline*, 5(1), pp. 49-50.

Patton, M. Q., 2002. *Qualitative research and evaluation methods*. 3rd ed. Thousand Oaks, CA: Sage.

Pollack, B. L., 2009. Linking the hierarchical service quality model to customer satisfaction and loyalty. *Journal of Services Marketing*, 23(1), pp.42-50.

Radnor, Z. J. and Barnes, D., 2007. Historical analysis of performance measurement and management in operations management. *International Journal of Productivity and Performance Management*. 56 (5-6), pp.384- 396.

Restaurant Business, 1989. Mystery shoppers: many operators bring problems out into the open by going under cover. 88(3), pp. 38-39.

Rubel, C., 1995. CSR works best when it also satisfies company needs. *Marketing News TM*, p. 38.

Rust, R. T., and Oliver, R. L., 1994. Service quality: Insights and managerial implications from the frontier. In: R. T. Rust, & R. L. Oliver, eds. *Service quality: Newdirections in theory and practice*. Thousand Oaks, CA: Sage. pp.1-19.

Sainy, R., 2010. A study of the effect of service quality on customer loyalty in retail outlets. Vilakshan: *The XIMB Journal of Management*, 7(2), pp. 49–62.

Saunders, M., Lewis, P. and Thornhill, A., 2009. *Research Methods for Business Students*. 5th ed. Essex: Pearson Education Limited.

Schwartz, M. S., and Schwartz, C. G., 1955. Problems in Participant Observation. *American Journal of Sociology*, 60(4), pp. 343-353.

Simons, R., 1991. Strategic Orientation and Top Management Attention to ControlSystems. *Strategic Management Journal*, 12, pp.49-62.

Smith, P., 1995. On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration,*18, pp.277-310.

Smith, J., Flowers, P. and Larkin, M., 2009. *Interpretative Phenomenological Analysis* (1st edn.). London: Sage Publications Ltd

Sobel, S. and Hines, G., 1990. Cray's new focus on customers. *Personnel Journal*, 69, pp. 59-63.

Spooner, L., 1985. Mystery shoppers uncover keys to sales and service. *Savings Institutions*, p. 92.

Strawderman, L. and Koubek, R., 2008. Human factors and usability in service quality measurement. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 18(4), pp.454-463.

Sureshchandar, G. S., Rajendran, C. and Kamalanabhan, T. J., 2001. Customer perceptions of service quality: A critique. *Total Quality Management*, 12(1), pp.111-124.

Turner, H., 2007. *Mystery shopping*. In: Market Research Handbook, ed. England: John Wiley & Sons Ltd. pp.333-347.

Ukko, J., Tenhunen, J. and Rantanen, H., 2007. Performance measurement impacts on management and leadership; perspectives of management and employees. *International Journal of Production Economics,* 110, pp.39-51.

Urban, W., 2013. Perceived Quality versus Quality of Processes: A Meta Concept of Service Quality Measurement. *Service Industries Journal*, 33(2), pp. 200-217.

Van der Wiele, T., Hesselink, M. and Van Iwaarden, J., 2005. Mystery Shopping: A tool to develop insight into customer service provision. *Total Quality Management & Business Excellence*, 16 (4), pp.529-541.

Varley, R. and Rafiq, M., 2004. *Principles of Retail Management.* Palgrave Macmillan: Basingstoke.

Vazquez, I. A., Rodriguez-del Bosque, A., Diaz, A. M., and Ruiz, A. V., 2001. Service Quality in Supermarket Retailing: Identifying Critical Service Experiences. *Journal of Retailing and Consumer Services*, 8(1), pp.1-14.

Waa, De. A., 2003. Behavioural factors important for the successful implementation and use of performance management systems. *Management Decision.* 41(8), pp. 688-697.

Wagnerová, I. and Baarová, E., 2008. *Mystery shopping*. In: Psychologie vekonomické praxi. Praha: Univerzita Karlova v Prahe, 41(1-2), pp. 8-10.

Webb, A. R., 2004. Managers' commitment to goals contained in a strategic performance measurement system. *Contemporary Accounting Research.* 21(4). pp.925-958.

Wilson, A. M., 1998. The use of mystery shopping in the measurement of service delivery. *The Service Industries Journal*, 18, pp.148-163.

Wilson, A. M., 2000. The use of performance information in the management of service delivery. *Marketing Intelligence and Planning*, 18(3), pp.127-134.

Wilson, A. M., 2001. Mystery Shopping: Using Deception to Measure Service Performance. *Psychology and Marketing*, 18(7), pp. 721-734.

Yin, R. K., 2011. *Qualitative research from start to finish.* New York, NY, US: Guilford Press.

Zeithaml, V.A., Berry, L. L., and Parasuraman, A., 1990. *Delivering Quality Service:Balancing CustomerPerception and Expectations*.New York: Free Press.

Zeithaml, V. A., Berry, L. L., and Parasuraman, A., 1996. The behavioral consequences of service quality. *The Journal of Marketing*, pp. 31–46.