

Mapping customer experience through mystery shopping: A case study in automotive dealerships

Abstract

Purpose: This study investigates mystery shopping as a diagnostic and strategic tool for assessing customer experience across the automotive retail customer journey.

Design/methodology/approach: The research adopts a case study methodology, using three sequential mystery shopper evaluations across a car-brand dealership network. It applies a structured assessment model across pre-visit, visit, and post-visit stages.

Findings: Service-quality scores generally improved over time, with significant gains observed in needs identification, vehicle presentation and overall score. Quartile analyses showed that high-performing dealerships consistently outperformed lower performers, while top improvers, though initially lower-scoring, demonstrated substantial gains over time. Comparisons with real Net Promoter Score data indicated that mystery shopping assessments became increasingly aligned with customer-reported outcomes, reflecting improved calibration across rounds.

Research limitations/implications: Limitations include the absence of purchase intention assessment and limited scoring scale granularity. Future research should expand the framework and integrate other evaluative tools.

Practical implications: Mystery shopping provides actionable insights for sales training, customer experience strategy, and operational consistency across dealerships.

Originality/value: This study introduces a structured, longitudinal mystery shopping framework that captures customer experience across time and touchpoints, extending beyond traditional one-off or compliance-focused applications. It contributes theoretically by operationalizing customer

experience as an observer-based, journey-centric approach, complementing self-reported service quality assessments and offering a repeatable method for longitudinal CX evaluation.

Keywords: Mystery shopping, Customer experience, Automotive retail, Service quality, CX evaluation, Dealership management

1. Introduction

Retailers compete in fragmented omnichannel environments (Reinartz et al., 2019; Shankar et al., 2021), where digital platforms reshape how customers search, compare, and engage with brands (Pantano et al., 2018; Reinartz et al., 2019). In such contexts, customer experience (CX) becomes a key strategic differentiator, especially in high-involvement sectors, such as automotive retail, where decisions span multiple channels and extended timeframes (Becker and Jaakkola, 2020; Bolton *et al.*, 2018).

Digital platforms intensify CX's relevance as platform dynamics change the nature of customer intelligence and require firms to continuously sense and adapt to the evolving customer journey (Castell *et al.*, 2023). However, while digital touchpoints play a crucial role in information search and early engagement, physical environments remain pivotal at critical decision-making moments, enabling interpersonal interaction, product evaluation, and purchase finalization. For retailers, misalignments between online interactions, in-store service, and strategic standards can undermine customer trust and perceived value. As established by the service-quality gaps model (Homburg *et al.*, 2017; Parasuraman *et al.*, 1985), the discrepancy between management perceptions of consumer expectations and the actual service delivery specifications can lead to fragmented CXs. Consequently, managing the digital–physical interface has become a key operational concern within contemporary retail and distribution networks.

Despite the growing emphasis on omnichannel CX, many existing approaches to service quality assessment remain limited in their ability to support this managerial challenge. Predominant measurement models rely heavily on customer self-reported perceptions or channel-specific performance indicators, offering only partial insight into how service quality is enacted across the customer journey (Manser Payne *et al.*, 2017; Patti *et al.*, 2020). Such approaches are often static and retrospective, providing limited diagnostic value for identifying where and how service inconsistencies emerge across fragmented touchpoints. Recent research underscores the importance of interaction-level and journey-spanning insights, suggesting a growing need for more operationally grounded evaluation methods (Dhaigude and Mohan, 2023; Singh *et al.*, 2021).

Unlike self-reported surveys, which suffer from recall bias (Patti *et al.*, 2020), mystery shopping (MS) provides a more objective audit of service standards. It bridges the gap between 'perceived quality' and 'actual service delivery' by employing trained observers who document specific behavioural markers in real-time (Jacob *et al.*, 2018; Rancati *et al.*, 2024). While traditionally employed for compliance monitoring, the scope of MS has gradually expanded to encompass strategic applications, such as CX benchmarking, salesforce training diagnostics, and continuous service enhancement (Blessing and Natter, 2019; Mattsson, 2012). In automotive retail, the physical showroom remains indispensable at the decision-making stage, where customers need direct interaction and test drives (ICDP, 2025). This, combined with the need to finalize the purchase and return for maintenance, means that measuring the in-person experience remains fundamentally relevant. Consequently, its potential to support the systematic evaluation and alignment of service quality across digital and physical channels over time remains underexplored in retail and distribution research.

A gap persists in the literature regarding how retailers can operationally evaluate and align service quality across fragmented customer journeys. Specifically, limited empirical guidance exists on frameworks that integrate digital and physical touchpoints and translate observational insights into action. To address this gap, the present study applies a structured, longitudinal, journey-centric MS framework within an automotive retail and distribution context. Accordingly, the study addresses the following research question: How can retailers and distributors use MS to effectively evaluate and align service quality across fragmented digital and physical touchpoints of the CX? Unlike prior studies that adopt static service quality models or focus on isolated customer journey moments, this research introduces a hybrid approach that integrates observational data across digital and physical touchpoints throughout the temporal evolution of CX delivery. It traces dealership performance across key stages of the customer journey: pre-visit digital engagement, in-store sales experience, and post-visit follow-up through three successive rounds of mystery shopper visits.

Through a cyclical assessment-and-feedback process, the study establishes a repeatable framework designed to enable continuous quality refinement. The program's transparent, iterative evaluations and adherence to established best practices enable it to function not only as a diagnostic instrument but also as a strategic support mechanism, fostering attentiveness to service quality and promoting more consistent delivery across touchpoints.

The research's contribution is threefold. Theoretically, it contributes to retail and distribution management literature by advancing a behaviour-based, journey-level approach to service quality evaluation. By clarifying what MS can capture (observable frontline behaviours) and what it cannot (e.g., customer decision-making processes), the research delineates the method's conceptual boundaries within retail service measurement. Methodologically, the study proposes a

scalable, empirically grounded approach for evaluating service performance across customer journey phases, enhancing the analytical sophistication of MS-based assessments. Managerially, the findings provide actionable guidance for dealership management, salesforce training, and CX innovation by demonstrating how a structured, cyclical evaluation process can support service consistency and continuous improvement. Its replicable framework positions it as a valuable reference for retailers aiming to close persistent gaps between service design and service delivery in increasingly customer-centric environments.

2. Research Background

2.1. Defining Customer Experience and Service Quality in Retail

Customer experience is a holistic interactive process, that involves customer's feelings, perceptions and attitudes developed throughout the successive stages of consumption (Jain *et al.*, 2017; Lemon and Verhoef, 2016; McColl-Kennedy *et al.*, 2019). This process-oriented view recognises CX as a dynamic and cumulative phenomenon, influenced by emotional, digital, relational, and operational touchpoints (Bolton *et al.*, 2018; Verhoef *et al.*, 2009). Store design, atmosphere, staff engagement, and transactional efficiency shape this experience (Bascur and Rusu, 2020; Berman, 2019; Bruckberger *et al.*, 2023; Mohd-Ramly and Omar, 2017). The role of staff, namely sales advisors that act as brand ambassadors is crucial (Cardell and Batra, 2023). Their responsiveness and empathy affect customer trust and decision-making (Blessing and Natter, 2019). Operational disruptions, such as poor communication or lack of personalisation, lead to negative experiences (Grewal and Roggeveen, 2020; Stein and Ramaseshan, 2016), while positive and emotionally resonant interactions contribute to loyalty and advocacy (Grönroos and Voima, 2013). Further, increasingly, technological advances and digitalization emerge as contributors to

brand-customer interaction opportunities (Bolton *et al.*, 2018). Today's customers navigate complex ecosystems of online reviews, showroom visits, and post-sale services, pressuring the alignment across these environments to deliver a coherent and trusted experience (Gao and Jiang, 2024).

This is particularly true in the retail sector, which has been facing changes over the last decades (Faria *et al.*, 2022; Reed *et al.*, 2004). As retail environments become increasingly saturated and consumer expectations evolve (Shankar *et al.*, 2021), service quality has assumed a central role in maintaining competitiveness and brand differentiation (Faria *et al.*, 2022). Extensive research highlights that superior service delivery not only satisfies but often exceeds customer expectations, leading to enhanced loyalty and long-term profitability (Bagherzadeh *et al.*, 2020; Olsson *et al.*, 2022). In retail sectors where the customer journey involves complex decision-making and high-value transactions, service quality assumes a central role (Bolton *et al.*, 2022). More than being a differentiator, it is a baseline expectation for sustaining market position in a fragmented and competitive landscape (Frasquet *et al.*, 2017).

2.2. Methodological Approaches to Service Quality Evaluation

Service quality evaluation covers diverse methodological approaches, each offering distinct insights into CX. The SERVQUAL framework (Parasuraman *et al.*, 1985, 1988) assesses service quality across five dimensions: tangibles, reliability, responsiveness, assurance, and empathy. SERVPERF (Cronin and Taylor, 1992) focuses solely on performance outcomes, positing that service execution drives customer satisfaction. Further methodologies, such as the customer effort score (Dixon *et al.*, 2010), quantify the ease with which customers complete interactions, recognizing that lower effort levels correlate with higher retention and loyalty. Additionally, the

expansion of digital platforms highlighted the value of online reviews and social media analysis in capturing real-time customer sentiment and service insights (Kannan and Li, 2017).

Despite their theoretical value, these models' reliance on self-reporting makes them susceptible to recall and social desirability biases (Rosenman et al., 2011). They also struggle to reflect the full complexity of real-world service interactions, especially those that unfold over multiple channels and time periods. In contrast, MS provides a structured and observational methodology that allows for consistent, real-time assessment of frontline service delivery (Wilson, 1998). MS enables objective evaluation of service delivery through three dimensions: 1) employee behaviour, 2) process compliance, and 3) experiential quality (e.g., customer's direct experience and the tangible output of the service encounter) (Douglas, 2015; Wilson, 1998; 2001). Traditionally employed as a compliance check, MS evolved into a strategic tool for capturing CX across multiple touchpoints and a performance management mechanism, providing insights to training, benchmarking, and customer journey optimisation (Wilson, 2001). Unlike customer satisfaction surveys, MS offers the advantage of observer-based evaluation, minimizing subjectivity and may enable a more consistent measurement of service encounters across time and location (Finn and Kayandé, 1999). It supports both formative feedback and summative performance tracking, providing actionable data for targeted interventions (Bettencourt *et al.*, 2013; Porter and Heyman, 2018). This not only allows structured scenario-based insights into how effectively service processes align with customer expectations but also shows process adherence and empowers managers to identify operational blind spots and design service improvements (Ford *et al.*, 2011; Porter and Heyman, 2018).

2.3. The Application of Mystery Shopping in Automotive Retail

MS holds particular value in automotive retail, a context marked by high involvement, long decision cycles, and complex interactions (Hashem et al., 2024). Although recent developments reflect evolving consumer preferences for convenience and transparency, the process remains comparatively less digitalized than other retail sectors (ICDP, 2025). In this research, the MS framework is conceptually anchored in the literature on CX (e.g., De Keyser *et al.*, 2020; Lemon and Verhoef, 2016) and service encounters (e.g., Bolton *et al.*, 2022; Van Doorn and Verhoef, 2008). As shown in Figure 1, the assessed touchpoints correspond to key journey stages. Evaluating these interlinked stages gives MS a longitudinal, 360-degree CX perspective, enabling gap detection, improvement monitoring, and alignment of operations with customer expectations.

Insert figure 1.

The pre-visit phase constitutes a critical stage in the customer journey, as it establishes the foundation for subsequent perceptions and behaviours. While physical touchpoints can occur, this stage often begins through digital touchpoints, such as websites, social media and online review platforms, where customers first encounter and evaluate a brand. At this stage, expectations are shaped largely by the brand's responsiveness, accessibility and coherence across channels. Research indicates that consumer-generated content and electronic word-of-mouth have a greater influence on customers than traditional advertising methods, highlighting the strategic importance of cultivating a credible and engaging online presence (Gensler *et al.*, 2013).

As the pre-sale often begins digitally, the usability of the interface and the quality of interaction with conversational agents become key determinants of initial service quality (Verhoef *et al.*, 2009; Wu *et al.*, 2023). The effectiveness of these technology-empowered interactions is a crucial

frontline factor in shaping customer perceptions and facilitating the pre-sale process (Marinova *et al.*, 2017). As a primary touchpoint in the customer journey, the pre-visit inquiry requires high functional and relational quality. Response time and personalization are thus critical indicators of effectiveness; prompt, relevant replies signal attentiveness and build relational capital, while delays or impersonal responses can constitute a critical incident that damages the customer-brand relationship (Van Doorn and Verhoef, 2008). MS evaluations at this stage assess dealerships' ability to engage leads promptly and professionally, aligning with research that highlights the importance of first impressions in the decision-making process (Kannan and Li, 2017; Wilson, 1998).

The physical dealership visit is a critical touchpoint, as sensory interaction with the vehicle fosters deeper emotional engagement than remote interactions (Lemon and Verhoef, 2016). For instance, the ability of representatives to assess customer needs, provide relevant product knowledge, and offer tailored recommendations may assist the customer's shopping experience (Pappas *et al.*, 2023; Wilder *et al.*, 2014), whereas inefficiencies such as long wait times, system errors, or inattentive behaviour can hamper it (Berry *et al.*, 2006; Bolton *et al.*, 2022). Contemporary CX theory conceptualizes the customer journey as a dynamic structure of interconnected touchpoints that must be deliberately orchestrated to drive satisfaction (Becker and Jaakkola, 2020). This orchestration requires ensuring that each touchpoint, from the initial greeting to the test drive, is effectively executed. As De Keyser *et al.* (2020) articulate, customers evaluate these moments across key dimensions; therefore, delivering a coherent and seamless overall experience hinges on excelling in critical human qualities like professionalism and empathy, and functional qualities, such as process clarity.

Lastly, the post-visit phase, comprising follow-up calls, after-sales services, complaint resolution or CRM outreach, represents a critical component of relationship marketing in automotive retail that supports the transition from one-off transactions into ongoing customer relationships (Lemon and Verhoef, 2016). Research suggests that businesses frequently underestimate the value of these interactions in cultivating long-term loyalty (Newbery and Farnham, 2013). Nevertheless, proactive outreach, meaningful contact, and empathetic communication have been shown to enhance satisfaction, strengthen trust and foster positive brand advocacy (e.g., Bhati and Verma, 2020), whereas poor customer support or rigid processes can erode trust and increase attrition (e.g., Bolton *et al.*, 2022). By integrating behavioural touchpoints with relational and emotional dimensions, this phase exemplifies how organizations can co-create value with customers and manage the cumulative experience across the customer journey (Verhoef *et al.*, 2009; Wu *et al.*, 2023). Through structured observations, MS offers a way to assess the consistency, timing, and personalization of post-visit contact, offering insights into customer retention and loyalty mechanisms (Healy *et al.*, 2007; Zubanov *et al.*, 2023).

3. Methodology

3.1. Research Design

The research adopts a quantitative case study design. A case study, understood as an in-depth, contextual examination of a bounded system, allows detailed exploration of service interactions in their natural environment. This approach is consistent with the aim of capturing nuanced, real-world behaviours and customer-facing processes (Yin, 2018). Data was collected from a South Korean automotive brand operating in Portugal, a mature "test-bed" market that enables near-census-level evaluation of the national network. The Portuguese market anticipates broader

European trends, with a 21-22% battery-electric vehicle share (ACAP, 2025) compared to Europe's 17.4% (ACEA, 2025). South Korean brands have grown to approximately 8% of the European market (ACEA, 2025; JATO Dynamics, 2025), making dealership service a critical differentiator. The focal brand, best known for its SUV portfolio, positions itself as an innovative value-for-money competitor and maintains a geographically well-distributed retail and after-sales network across Portugal. Ranked among the ten best-selling automotive brands nationally, it offers an appropriate empirical setting for examining premium CX delivery in a competitive transition context.

The study analyses the brand's full Portuguese dealership network, comprising 33 dealerships, and examines the evolution of the customer journey across sales and after-sales touchpoints over time. The customer journey was assessed considering three key stages, weighted according to the brand's protocol: pre-visit (20% weighting), visit (70%), and post-visit (10%). These weights reflect the strategic emphasis placed by the brand on the in-store experience as the most influential in shaping customer decisions, based on both internal CX research and industry benchmarks. The pre- and post-visit phases, while critical, were considered supportive of the central sales interaction and thus received proportionally lower weightings. The journey stages analysed included both digital and physical touchpoints.

3.2. Instrument Design

The MS evaluation framework is grounded in established service quality models, such as SERVQUAL, and in industry benchmarks refined through more than 20 years of market experience. Each evaluation used a standardized observation form, developed in collaboration with the organization's quality assurance team, and applied ten service quality assessment criteria

reflecting critical CX touchpoints of the involved stages. The instrument focused on observable behaviours rather than subjective perceptions (Wilson, 1998). This foundation provided a basis for calibrating key service touchpoints and assigning quantitative scores to specific components of the CX. Qualitative experiences were systematically translated into measurable performance data, allowing for objective assessment of service quality and clear identification of trends and areas for improvement (Wilson, 1998). This was achieved through the operationalization of service standards into binary (Yes/No) and anchored Likert-type scales. Each qualitative dimension was decomposed into specific, observable behavioural markers, such as the use of a formal greeting or the provision of a business card. This minimized observer subjectivity and captured qualitative interactions as objective data points for standardized assessment. Following the framework for measuring service performance through structured observation (Wilson, 2001), the assessment was categorized into three distinct stages, with ten service quality assessment criteria used. The pre-visit stage measured lead response efficiency, including response time, and quality (which considered message clarity, personalization, and invitation to visit the dealership). The visit stage captured welcoming/receptiveness, through initial greeting/approach; customer needs identification; vehicle presentation; test drive execution; sales process, encompassing the explanation of financing options, warranty conditions, and commercial proposals; trade-in discussion; and farewell etiquette. The post-visit phase assessed the presence, quality, and timeliness of follow-up communication. “Did the sales consultant have a neat appearance?”, “Did the sales consultant summarize all your needs and validate them with you”, “Did the sales consultant contact you within 3 days after the visit?” are sample questions for each phase.

3.3. Data Collection

Data was collected using multiple mystery shoppers, selected from a database of highly trained professionals experienced in conducting such surveys within the industry. They posed as real customers across both digital and physical touchpoints, submitting online enquiries, visiting showrooms, engaging in test drives, and observing post-visit communication, reflecting omnichannel service experiences.

Before data collection, mystery shoppers completed structured training that included orientation on the evaluation instrument, calibration through pilot scenarios, and inter-rater reliability testing to ensure consistent interpretation. During data collection, they followed a scripted interaction briefing, ensuring consistency and comparability across all observations. To maintain objectivity, no shopper evaluated the same dealership more than once, even though the shopper could visit others, for consistency issues. The collection occurred through three rounds of MS assessments conducted over a defined time horizon to capture dealership performance progression: end of December 2023 (round 1), end of June 2024 (round 2), and end of December 2024 (round 3), consisting of a total of 990 observations ($33 \text{ dealerships} \times 3 \text{ rounds} \times 10 \text{ service quality criteria}$). This timeframe enabled homologous period comparisons and accounted for seasonal market fluctuations. After each assessment round, dealerships received aggregated feedback in a report that included summary statistics, benchmark comparisons and anonymized qualitative comments. Based on this, the retailer was able to adjust training and service protocols between assessment rounds.

4. Results

The scoring model applied weighted criteria across the three journey phases, producing a composite performance score per dealership per evaluation cycle. Scores were aggregated and

analysed both longitudinally and cross-sectionally, with the scores from 90 to 100 being considered “excellent”. Table I presents the results.

Insert table I.

Descriptive analyses of service-quality scores across the three assessment periods indicated general improvements, though the magnitude and timing varied by dimension. To test the statistical significance of these changes, we conducted one-way ANOVAs. Given potential deviations from normality, we used bootstrap post-hoc comparisons (5,000 samples; Field, 2013). This approach ensures that the findings are not artifacts of the data distribution and provides more reliable bias-corrected confidence intervals. Most dimensions did not show significant changes; however, the needs identification criterion indicated consistent gains after round 1 (round 1 vs. 2 BCa 95% CI: 0.64 to 4.00 and round 1 vs. 3 BCa 95% CI: 0.64 to 4.00). Significant changes also emerged in vehicle presentation, with round 2 outperforming round 1 (BCa 95% CI: 0.00 to 7.35) and round 3 (BCa 95% CI: -10.29 to -0.00). Further, the overall score improved from round 1 to round 3 (BCa 95% CI: -9.83 to -0.54). Results indicate dealership improvements were selective, not systemic. Gains in needs identification suggest codified diagnostic behaviours respond well to MS, while temporary vehicle presentation gains highlight the challenge of sustaining experiential behaviours. Aggregate scores masked this uneven adaptation, implying networks internalize process-driven aspects first. Thus, MS initiates targeted shifts but requires ongoing support for holistic change.

To understand patterns of improvement, participants were grouped into quartiles. The first analysis used overall final scores at round 3, dividing dealerships into the top and bottom quartiles to contrast high and low end-stage achievers. This approach provides a performance-oriented perspective, identifying which dimensions distinguish final achievement levels and offering insight into the competency profile associated with superior outcomes. Independent-samples comparisons across all criteria revealed clear patterns of differentiation. High achievers consistently outperformed the bottom quartile across multiple dimensions, with the largest gaps observed in response time, response quality and test drive at round 3 (mean difference=68.75, 45.00 and 43.50, Cohen's d =2.96, 2.06 and 2.20, respectively). These results indicate that certain operational or service processes were significantly better implemented by top performers. Other dimensions, including welcoming, vehicle presentation, sales process, trade in, farewell and follow up also showed significant differences, with moderate to large effect sizes, suggesting that high-performing firms maintained consistent quality across multiple service dimensions.

The second analysis was based on total change from round 1 to 3, contrasting the top 25% of improvers with the bottom 25%. This approach highlights differential gains, identifying which participants benefited most from the assessment process irrespective of their initial performance. At round 1, the top improvers showed a lower score than the bottom improvers (74.67 vs 90.24), a statistically significant difference with a large effect size (Cohen's d =2.18). Nevertheless, these dealerships seemed to leverage the iterative feedback, improving in most dimensions, particularly response time, response quality and test drive. These findings emphasize the value of MS assessments both in identifying high-performing practices and in revealing areas where targeted interventions may be necessary for lower-performing firms.

In a supplementary test, we accessed the brand's real net promoter score (NPS) data for the analysed periods and compared it with MS assessments to evaluate convergence, improvement, calibration, and growth. NPS scores were 92.90, 94.00, and 86.70, indicating that MS assessments underestimated in the first rounds, and slightly overestimate in the later round. A repeated-measures ANOVA on the absolute differences between MS and NPS scores revealed a significant effect of time, $F(2, 64)=16.46$, $p<0.001$, $\eta^2=0.34$, with pairwise comparisons showing reduced discrepancies from rounds 1 and 2 to round 3, reflecting improved calibration over time. Nevertheless, a paired-samples t-test comparing MS assessment growth ($M=11.30$, $SD=13.77$) to NPS growth ($M=0.00$, $SD=0.00$) showed that the first round significantly exceeded actual performance, $t(32)=4.71$, $p<0.001$, $d=0.82$. These findings suggest that while MS assessments became more aligned with objective performance, they continued to overestimate performance compared to actual NPS growth, highlighting persistent optimism despite improved calibration.

5. Discussion

These findings underscore MS's strategic value for evaluating and enhancing CX in automotive retail. By conducting three sequential waves of evaluation, the study provides a longitudinal lens through which service performance can be tracked, revealing both measurable improvements and persistent operational gaps that challenge the consistency of CX delivery.

Our findings demonstrate that improvements in service performance are not merely an outcome of increased monitoring, but rather a reflection of broader organizational alignment processes. While Parasuraman et al. (1985) conceptualize service gaps between design and execution, our evidence shows structured evaluation can partially mitigate these gaps by reinforcing learning and

behavioural consistency. In this respect, the results support Homburg *et al.*'s (2017) assertion that CX constitutes a strategic and manageable construct, provided that organizations employ objective, behaviourally-grounded diagnostic tools capable of revealing how service is actually delivered at the frontline.

The data supports the initial proposition that structured MS can function as a feedback mechanism that enables the detection of service progression over time. However, the observed effects are uneven across the customer journey, suggesting that the influence of MS is contingent on the nature of the behaviours being evaluated. The cyclical evaluation process appears to reinforce learning and operational discipline, consistent with prior research indicating that structured feedback systems facilitate behavioural alignment and incremental refinement (Grönroos and Voima, 2013). Nevertheless, the selective pattern of improvement indicates that MS primarily strengthens practices that are procedural, teachable, and already embedded within existing operational routines. This aligns with literature suggesting that audit-based and evaluative mechanisms are most effective when the target behaviours are codifiable to be embedded in frontline scripts or checklists (e.g. Blessing and Natter, 2019).

In contrast, dimensions displaying modest or statistically inconclusive improvement point to the conceptual and practical limits of MS as a mechanism for influencing deeper behavioural or culturally-embedded aspects of CX. CX is inherently complex and relational, requiring situational interpretation, emotional regulation, and interpersonal adaptability capabilities (Lemon and Verhoef, 2016). These elements draw on tacit knowledge and broader service climate, factors shaped gradually within organizations and are, therefore, less responsive to evaluative pressure or behavioural monitoring alone (Bowen and Schneider, 2014). These gaps reflect the difficulty of

shifting relational competencies through measurement alone, without complementary cultural enablers (Kannan and Li, 2017; Verhoef *et al.*, 2009).

The asymmetry between standardized and variable touchpoints reinforces this interpretation. High and relatively stable performance in routine operational metrics suggests that dealerships can reliably embed procedural behaviours through system design and formalized routines. Conversely, touchpoints requiring interpersonal adaptability remain more fragile, reflecting the established distinction between codifiable reliability and relational value creation (Grönroos and Voima (2013). Variation across dealerships, thus, indicates underlying differences in capability configurations, differences that MS can identify but not, per se, resolve.

At the network level, the pattern of results indicates neither uniform convergence, nor immutable divergence, but differentiated learning trajectories. While some gaps between top- and bottom-performing dealerships persisted across rounds, suggesting stable structural differences, other dealerships, that began from weaker starting points, demonstrated disproportionately large improvements. This suggests that MS may interact with store-level absorptive capacity, rather than operating as an equalising mechanism (Blessing and Natter, 2019). Dealerships with stronger managerial coordination, or complementary resources may be better able to translate evaluative feedback into adjustments (Marinova *et al.*, 2017), whereas others may require broader organizational support to do so (Douglas, 2015). Although the study cannot infer specific action taken by individual dealerships following each assessment round, the observed variation is consistent with the notion that service quality improvement depends on how effectively organisations translate evaluative information into frontline behavioural change.

The comparison with real NPS data provides a complementary behavioural benchmark that further contextualises these findings. Although MS evaluates service execution rather than customer

perceptions, the progressive reduction in the MS–NPS discrepancy over time suggests increasing alignment between frontline service delivery and customer-reported loyalty intentions. This does not establish causality but strengthens the construct validity of the MS framework by indicating directional consistency between operational execution quality and experiential outcomes.

These findings also highlight the conceptual boundaries of MS within the broader CX measurement landscape. MS provides itemized, real-time behavioural observations, yet it does not capture customers' internal perceptions, expectations or emotional evaluations. Perceptual methods, such as SERVQUAL (Parasuraman *et al.*, 1985, 1988) and customer satisfaction surveys, are designed to access these experiential dimensions but remain subject to biases, such as selective memory, social desirability, and misinterpretation. The partial alignment between MS results and NPS scores illustrates the well-established distinction between observed behaviour and lived experience (Klaus and Maklan, 2013). Integrating MS with perceptual indicators, therefore, provides a more complete depiction of CX, enabling organizations to examine how operational execution translates into value-in-use for customers across fragmented digital and physical touchpoints.

6. Conclusion and Implications

This research offers new insights into the evolving field of CX assessment by demonstrating the application of MS as a viable, structured approach for capturing real-time service quality across multiple stages of the customer journey. While existing literature highlights the multi-dimensional nature of CX (e.g., Lemon and Verhoef, 2016), few studies have operationalised this concept using observational methods that bridge the gap between theory and practice. The present study employs MS in that regard, providing useful theoretical and managerial implications.

6.1. Theoretical implications

Theoretical implications include reinforcing that service quality orchestration is not a static achievement, but a dynamic process. Therefore, CX governance can benefit from systematic feedback loops (Homburg *et al.*, 2017), especially in high involvement retail contexts that involve distributors and franchised retailers. This research contributes to the development of hybrid CX evaluation models by demonstrating how MS can be a reliable empirical layer atop traditional survey-based tools, thus enabling a more holistic and behaviourally-grounded assessment of service delivery across time and channels. While established models, such as SERVQUAL (Parasuraman *et al.*, 1988) and SERVPERF (Cronin and Taylor, 1992), have enriched theoretical understandings of service quality, their dependence on customer self-reporting renders them vulnerable to recall bias, framing effects, and inconsistent interpretations. Conversely, the MS approach offers a behaviourally-grounded, replicable method that aligns with marketing research calls to bridge conceptual rigour and real-world relevance (Foltean, 2019). By structuring the protocol around customer journey stages, the study demonstrates a hybrid approach that advances theory while addressing practical service performance challenges. Extending the work of Wilson (1998), the study contributes to both service marketing and digital marketing literatures as it empirically validates that, despite the critical importance of digital touchpoints in the modern customer journey, in some retail sales contexts, the physical retail environment remains an indispensable touchpoint at key decision-making and post-visit stages. Our longitudinal MS data confirms that the in-dealership experience, where customers seek final consultation, test drives, and transactional closure, is a primary determinant of service quality perception. Furthermore, the post-sale visit for maintenance and assistance solidifies the customer relationship. Consequently,

this research ratifies the enduring relevance of evaluating physical interactions through MS, affirming its critical role in a comprehensive, omnichannel CX strategy

6.2. Managerial implications

The findings of this study translate into several actionable implications for dealership management and automotive retail strategy. Rather than a compliance tool, managers can leverage MS for continuous improvement. While MS alone does not drive behavioural or cultural change, it generates actionable visibility around service execution, enabling organisations to prioritise resources, tailor developmental efforts, and monitor progress over time. This requires building integrated feedback processes. A critical first step is to institutionalize the MS feedback cycle by formally scheduling post-round review sessions with sales teams to deconstruct the results, focusing on translating specific, low-scoring behaviours, such as passive greetings, into role-playing scenarios and targeted coaching sessions. By making this data a regular part of team meetings and individual performance coaching, MS becomes a tool for professional development rather than punitive measures, thereby fostering a culture of accountability and continuous learning. Furthermore, the findings demonstrate that not all service gaps are created equal, underscoring the need for managers to diagnose the underlying drivers of underperformance rather than rely on uniform corrective actions. Because some dimensions respond more readily to structured evaluation while others remain resistant, managers should apply tailored responses that reflect whether the gap is procedural, relational, or capability-based. For trainable, procedural gaps, like sales process clarity, or follow-up protocols, MS data provides a blueprint for developing standardized operating procedures and training modules. However, for more entrenched behavioural and cultural gaps, such as initial greeting enthusiasm, training alone is insufficient and

must be complemented by initiatives that build intrinsic motivation, and a customer-centric culture that includes empowering employees, recognizing exemplary behaviours, and hiring for empathy. Most critically, some systemic gaps (e.g., response time) may not be addressed through training alone, suggesting the need for strategic consideration of underlying technologies and workflows. Consistent with prior research on marketing automation and AI-enabled systems (Davenport *et al.*, 2020), MS programs may offer a business case for investments in Customer Relationship Management systems and automated response mechanisms to support timely initial contact. The persistent weakness in responsiveness-related dimensions, particularly when contrasted with stronger performance in face-to-face procedural elements, suggests a potential misalignment between digital or system-based capabilities and frontline execution. This pattern indicates that senior leadership may need to prioritise investments that better integrate digital and physical components of the service journey. Finally, the longitudinal nature of the MS framework means that managers can use it to set progressively higher performance expectations over time and incorporate objective experience metrics into performance reviews or incentive systems. This enables rewarding both sales outcomes and CX quality (Blessing and Natter, 2019), fostering loyalty and brand equity (Homburg *et al.*, 2017). By adopting this more strategic, integrated, and diagnostic approach, managers can transform their MS program into a key component of an integrated continuous improvement system.

6.3. Limitations and Future Research

The findings of this study should be interpreted considering its limitations. First, while the findings may offer insights for high-involvement retail contexts with in-person interactions, the single-case design in automotive retail limits generalisability. Future research should test this approach across

multiple industries and cultures, including within-industry differences such as premium versus mass-market segments.

Second, the MS operationalisation involves subjective weighting and scoring decisions. Although sensitivity analyses confirmed result stability, the weights reflect the brand's strategic priorities. Findings might shift with different touchpoint importance levels. Future research could use dynamic or customer-derived weights to refine diagnostic precision. Furthermore, the uniform scoring system may oversimplify context-dependent encounters; triangulating with qualitative feedback could capture richer dimensions.

Third, while data collection across three time points allowed comparison, the temporal scope is insufficient for assessing long-term trends and causality. Purchase intention was not measured, limiting links to commercial outcomes. Future research should extend observation periods and integrate behavioural data.

Finally, the findings focus on systematic behavioural indicators, not the full experiential context. Future research should combine MS with customer-reported measures, like interviews, for a more holistic understanding.

References

- ACAP–Associação Automóvel de Portugal (2025), “Estatísticas do sector automóvel”, available at [acap.pt](https://www.acap.pt) (accessed 11 November 2025).
- ACEA–European Automobile Manufacturers Association (2025), “New car registrations”, available at www.acea.auto (accessed 8 May 2026).

- Bagherzadeh, R., Rawal, M., Wei, S. and Torres, J.L.S. (2020) “The journey from customer participation in service failure to co-creation in service recovery”, *Journal of Retailing and Consumer Services*, Vol.54,pp.102058.
- Bascur, C. and Rusu, C. (2020), “Customer experience in retail: A systematic literature review”, *Applied Sciences*, Vol.10 No.21,pp.7644.
- Bhati, R. and Verma, H.V. (2020), “Antecedents of customer brand advocacy: A meta-analysis of the empirical evidence”, *Journal of Research in Interactive Marketing*, Vol.14 No.2,pp.153-172.
- Becker, L. and Jaakkola, E. (2020), “Customer experience: Fundamental premises and implications for research”, *Journal of the Academy of Marketing Science*, Vol.48 No.4,pp.630-648.
- Berman, B. (2019), “Flatlined: Combatting the death of retail stores”, *Business Horizons*, Vol.62 No.1,pp.75-82.
- Berry, L.L., Wall, E.A. and Carbone, L.P. (2006), “Service clues and customer assessment of the service experience: Lessons from marketing”, *Academy of Management Perspectives*, Vol.20 No.2,pp.43-57.
- Bettencourt, L.A., Brown, S.W. and Sirianni, N.J. (2013), “The secret to true service innovation”, *Business Horizons*, Vol.56 No.1,pp.13-22.
- Blessing, G. and Natter, M. (2019), “Do mystery shoppers really predict customer satisfaction and sales performance?”, *Journal of Retailing*, Vol.95 No.3,pp.47-62.
- Bolton, R.N., Gustafsson, A., Tarasi, C.O. and Witell, L. (2022), “Designing satisfying service encounters: website versus store touchpoints”, *Journal of the Academy of Marketing Science*, Vol.50 No.1,pp.85-107.

- Bolton, R.N., McColl-Kennedy, J.R., Cheung, L., Gallan, A., Orsingher, C., Witell, L. and Zaki, M. (2018), "Customer experience challenges: bringing together digital, physical and social realms", *Journal of Service Management*, Vol.29 No.5, pp.776-808.
- Bowen, D.E. and Schneider, B. (2014), "A service climate synthesis and future research agenda", *Journal of Service Research*, Vol.17 No.1, pp.5-22.
- Bruckberger, G., Fuchs, C., Schreier, M. and Van Osselaer, S.M. (2023), "Retailing groundedness: How to improve customer experience, brand perceptions, and customer loyalty through feelings of groundedness", *Journal of Retailing*, Vol.99 No.4, pp.594-604.
- Cardell, M. and Batra, G. (2023), "Why human interaction remains essential to the car buying journey", 19 December, available at www.ey.com (accessed 23 July 2025).
- Castell, C., Kiefer, J., Schubach, S., Schumann, J.H., Graf-Vlachy, L. and König, A. (2023), "Integrating digital platform dynamics into customer orientation research: A systematic review and research agenda", *Journal of Business Research*, Vol.163, pp.113911.
- Chevalier, J.A. and Mayzlin, D. (2006), "The effect of word of mouth on sales: Online book reviews", *Journal of Marketing Research*, Vol.43 No.3, pp.345-354.
- Cronin, J.J. and Taylor, S.A. (1992), "Measuring service quality: A reexamination and extension", *Journal of Marketing*, Vol.56 No.3, pp.55-68.
- Davenport, T., Guha, A., Grewal, D. and Bressgott, T. (2020), "How artificial intelligence will change the future of marketing", *Journal of the Academy of Marketing Science*, Vol.48 No.1, pp.24-42.
- De Keyser, A., Verleye, K., Lemon, K.N., Keiningham, T.L. and Klaus, P. (2020), "Moving the customer experience field forward: introducing the touchpoints, context, qualities (TCQ) nomenclature", *Journal of Service Research*, Vol.23 No.4, pp.433-455.

- Dhaigude, S.A. and Mohan, B.C. (2023), “Customer experience in social commerce: A systematic literature review and research agenda”, *International Journal of Consumer Studies*, Vol.47 No.5,pp.1629-1668.
- Dixon, M., Freeman, K. and Toman, N. (2010), “Stop trying to delight your customers”, *Harvard Business Review*, Vol.88 No.7/8,pp.116-122.
- Douglas, J. (2015), “Mystery shoppers: an evaluation of their use in monitoring performance”, *The TQM Journal*, Vol.27 No.6,pp.705-715.
- Faria, S., Carvalho, J.M.S. and Vale, V.T. (2022), “Service quality and store design in retail competitiveness”, *International Journal of Retail & Distribution Management*, Vol.50 No.13,pp.184-199.
- Finn, A. and Kayandé, U. (1999), “Unmasking a phantom: A psychometric assessment of mystery shopping”, *Journal of Retailing*, Vol.75 No.2,pp.195-217.
- Foltean, F.S. (2019), “Bridging marketing theory-practice gap to enhance firm performance: Introduction to the special issue”, *Journal of Business Research*, Vol.104,pp.520-528.
- Ford, R., Latham, G. and Lennox, G. (2011), “Mystery shoppers: A new tool for coaching employee performance improvement”, *Organizational Dynamics*, Vol.40 No.2,pp.157-164.
- Frasquet, M., Mollá Descals, A. and Ruiz-Molina, M.E. (2017), “Understanding loyalty in multichannel retailing: The role of brand trust and brand attachment”, *International Journal of Retail & Distribution Management*, Vol.45 No.6,pp.608-625.
- Gao, W. and Jiang, N. (2024), “Demystifying the combined effect of consistency and seamlessness on the omnichannel customer experience: A polynomial regression analysis”, *Journal of Theoretical and Applied Electronic Commerce Research*, Vol.19 No.1,pp.232-248.

- Gensler, S., Völckner, F., Liu-Thompkins, Y. and Wiertz, C. (2013), “Managing brands in the social media environment”, *Journal of Interactive Marketing*, Vol.27 No.4, pp.242-256.
- Grewal, D. and Roggeveen, A.L. (2020), “Understanding retail experiences and customer journey management”, *Journal of Retailing*, Vol.96 No.1, pp.3-8.
- Grönroos, C. and Voima, P. (2013), “Critical service logic: Making sense of value creation and co-creation”, *Journal of the Academy of Marketing Science*, Vol.41 No.2, pp.133-150.
- Hashem, M., Ruiz, C. and Curras-Perez, R. (2024), “Understanding the dynamics of brand love in the automobile industry”, *Journal of Theoretical and Applied Electronic Commerce Research*, Vol.19 No.2, pp.1142-1163.
- Healy, M.J., Beverland, M.B., Oppewal, H. and Sands, S. (2007), “Understanding retail experiences-the case for ethnography”, *International Journal of Market Research*, Vol.49 No.6, pp.751-778.
- Homburg, C., Jozić, D. and Kuehn, C. (2017), “Customer experience management: toward implementing an evolving marketing concept”, *Journal of the Academy of Marketing Science*, Vol.45 No.3, pp.377-401.
- ICDP–International Car Distribution Programme (2025), “The changing face of Europe’s automotive retail”, available at www.calameo.com (accessed 11 November 2025).
- Jacob, S., Schiffino, N. and Biard, B. (2018), “The mystery shopper: A tool to measure public service delivery?”, *International Review of Administrative Sciences*, Vol.84 No.1, pp.164-184.
- Jain, R., Aagja, J. and Bagdare, S. (2017), “Customer experience—a review and research agenda”, *Journal of Service Theory and Practice*, Vol.27 No.3, pp.642-662.
- JATO Dynamics (2025), “European new car market reports and industry analysis”, available at www.jato.com (accessed 8 May 2026)

- Kannan, P.K. and Li, H. (2017), "Digital marketing: A framework, review and research agenda", *International Journal of Research in Marketing*, Vol.34 No.1,pp.22-45.
- Klaus, P.P. and Maklan, S. (2013), "Towards a better measure of customer experience", *International Journal of Market Research*, Vol.55 No.2,pp.227-246.
- Lemon, K.N. and Verhoef, P.C. (2016), "Understanding customer experience throughout the customer journey", *Journal of Marketing*, Vol.80 No.6,pp.69-96.
- Manser Payne, E., Peltier, J.W. and Barger, V.A. (2017), "Omni-channel marketing, integrated marketing communications and consumer engagement: A research agenda", *Journal of Research in Interactive Marketing*, Vol.11 No.2,pp.185-197.
- Marinova, D., de Ruyter, K., Huang, M.H., Meuter, M.L. and Challagalla, G. (2017), "Getting smart: Learning from technology-empowered frontline interactions", *Journal of Service Research*, Vol.20 No.1,pp.29-42.
- Mattsson, J. (2012), "Strategic insights from mystery shopping in B2B relationships", *Journal of Strategic Marketing*, Vol.20 No.4,pp.313-322.
- McColl-Kennedy, J.R., Zaki, M., Lemon, K.N., Urmetzer, F. and Neely, A. (2019), "Gaining customer experience insights that matter", *Journal of Service Research*, Vol.22 No.1,pp.8-26.
- Mohd-Ramly, S. and Omar, N.A. (2017), "Exploring the influence of store attributes on customer experience and customer engagement", *International Journal of Retail & Distribution Management*, Vol.45 No.11,pp.1138-1158.
- Newbery, P. and Farnham, K. (2013), *Experience design: A framework for integrating brand, experience, and value*, John Wiley & Sons.

- Olsson, J., Osman, M.C., Hellström, D. and Vakulenko, Y. (2022), “Customer expectations of unattended grocery delivery services: Mapping forms and determinants”, *International Journal of Retail & Distribution Management*, Vol.50 No.13,pp.1-16.
- Pantano, E., Priporas, C.V. and Dennis, C. (2018), “A new approach to retailing for successful competition in the new smart scenario”, *International Journal of Retail & Distribution Management*, Vol.46 No.3,pp.264-282.
- Pappas, A., Fumagalli, E., Rouziou, M. and Bolander, W. (2023), “More than machines: The role of the future retail salesperson in enhancing the customer experience”, *Journal of Retailing*, Vol.99 No.4,pp.518-531.
- Parasuraman, A., Zeithaml, V.A. and Berry, L.L. (1985), “A conceptual model of service quality and its implications for future research”, *Journal of Marketing*, Vol. 49 No. 4,pp.41-50.
- 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*, Vol.64 No.1,pp.12-40.
- Patti, C.H., van Dessel, M.M. and Hartley, S.W. (2020), “Reimagining customer service through journey mapping and measurement”, *European Journal of Marketing*, Vol.54 No.10,pp.2387-2417.
- Porter, M.C. and Heyman, J.E. (2018), “We’ve shopped before: Exploring instructions as an influence on mystery shopper reporting”, *Journal of Retailing and Consumer Services*, Vol.45,pp.12-20.
- Rancati, G., Nguyen, T.T.T., Fowler, D., Mauri, M. and Schultz, C.D. (2024), “Customer experience in coffee stores: A multidisciplinary Neuromarketing approach”, *Journal of Consumer Behaviour*, Vol.23 No.1,pp.243-259.

- Reed, G., Story, V. and Saker, J. (2004), "Information technology: Changing the face of automotive retailing?", *International Journal of Retail & Distribution Management*, Vol.32, No.1,pp.19-32.
- Reinartz, W., Wiegand, N. and Imschloss, M. (2019), "The impact of digital transformation on the retailing value chain", *International Journal of Research in Marketing*, Vol.36 No.3,pp.350-366.
- Rosenman, R., Tennekoon, V. and Hill, L.G. (2011), "Measuring bias in self-reported data", *International Journal of Behavioural and Healthcare Research*, Vol.2 No.4,pp.320-332.
- Shankar, V., Kalyanam, K., Setia, P., Golmohammadi, A., Tirunillai, S., Douglass, T., Hennessey, J., Bull, J.S. and Waddoups, R. (2021), "How technology is changing retail", *Journal of Retailing*, Vol.97 No.1,pp.13-27.
- Singh, J., Nambisan, S., Bridge, R.G. and Brock, J.K.U. (2021), "One-voice strategy for customer engagement", *Journal of Service Research*, Vol.24 No.1,pp.42-65.
- Stein, A. and Ramaseshan, B. (2016), "Towards the identification of customer experience touch point elements", *Journal of Retailing and Consumer Services*, Vol.30,pp.8-19.
- Van Doorn, J. and Verhoef, P.C. (2008), "Critical incidents and the impact of satisfaction on customer share", *Journal of Marketing*, Vol.72 No.4,pp.123-142.
- Verhoef, P.C., Lemon, K.N., Parasuraman, A., Roggeveen, A., Tsiros, M. and Schlesinger, L.A. (2009), "Customer experience creation: Determinants, dynamics and management strategies", *Journal of Retailing*, Vol.85 No.1,pp.31-41.
- Wilder, K.M., Collier, J.E. and Barnes, D.C. (2014), "Tailoring to customers' needs: Understanding how to promote an adaptive service experience with frontline employees", *Journal of Service Research*, Vol.17 No.4,pp.446-459.

- Wilson, A.M. (1998), “The use of mystery shopping in the measurement of service delivery”, *Service Industries Journal*, Vol.18 No.3,pp.148-163.
- Wilson, A.M. (2001), “Mystery shopping: Using deception to measure service performance”, *Psychology & Marketing*, Vol.18 No.7,pp.721-734.
- Wu, Z., Aw, E.C.X. and Chuah, S.H.W. (2023), “Webrooming as the new retail experience: a smart-shopping perspective”, *International Journal of Retail & Distribution Management*, Vol.51 No.6,pp.736-754.
- Yin, R.K. (2018), *Case study research and applications: Design and methods*, (6thed.). SAGE Publications.
- Zubanov, N., Heinz, M., Block, S. and Friebe, G. (2023), “Mystery shopping as a strategic management practice in multi-site service firms”, In *Academy of Management Proceedings*. Academy of Management, New York, Vol.2023 No.1,p.15600.