Assessing the impact of mystery client traits on service evaluation
Pedro Quelhas Brito, Meena Rambocas,

Article information:
Permanent link to this document: https://doi.org/10.1108/JSM-01-2015-0004

Users who downloaded this article also downloaded:

Access to this document was granted through an Emerald subscription provided by emerald-srm:363333 []

For Authors
If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldinsight.com/authors for more information.

About Emerald www.emeraldinsight.com
Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.
Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

*Related content and download information correct at time of download.
Assessing the impact of mystery client traits on service evaluation

Pedro Quelhas Brito
Faculdade de Economia, Universidade do Porto, Porto, Portugal, and
Meena Rambocas
Department of Management Studies, The University of the West Indies (St. Augustine Campus), St. Augustine, Trinidad and Tobago

Abstract
Purpose – This study aims to investigate the reliability of a mystery client (MC) as a service evaluation technique taking into consideration personal differences of the MC agents.

Design/methodology/approach – The ratings from 144 MCs from 355 evaluations of computer and electronic stores were cross analyzed with eight psychographic and demographic profile variables.

Findings – MCs who were highly involved in the product category were more critical of service responsiveness with respect to product demonstrations and listening to customer requirements. On the other hand, MCs with stronger faith in intuition were more inclined to rate services higher on empathy with respect to employees making a conscientious effort to understand customers’ needs.

Practical implications – Depending on the service marketing goals, managers learn to define which aspects of MC profile they should consider or avoid during the recruitment as well as becoming more critical when they analyze the evaluation reports to avoid an interpretation bias.

Originality/value – The usefulness of the MC tool relies on its reliability and credibility as a marketing research technique. It was identified that the MC personality traits are more likely associated with marketing service evaluation variability.

Keywords Reliability, Service evaluation, Personality traits, Mystery client

Paper type Research paper

1. Introduction

1.1 Mystery client concept
Mystery client (MC) survey is a data collection technique relevant to business service quality assessment and scientific research. Although impactful in the insights that it can improve service quality, the North American MSPA organization (2014) cautions that the MC survey is not exclusive or sufficient, but a complementary method to other research tools. In the scientific literature, MC has been applied in several service research contexts: health (Gosselt et al., 2007; Young et al., 2009; Glasier et al., 2010; Wong et al., 2012), banking (Tarantola et al., 2012), family planning (León et al., 1994), hospitality tourism and transportation (Nathanail, 2008; Butcher et al., 2009) and retailing (Finn and Kayandé, 1999; Gómez et al., 2011; Wang et al., 2012).

Table I combines several definitions of MC surveys. In marketing management where the service encounter is the main interface between a customer and service provider, the more general term “client” will be adopted here instead of “shopper” term. Although both terms overlap the same potential buyer function, the former emphasizes the service experience process, while the latter is focused more on the transaction. From these definitions, it was possible to depict the following common features of the MC concept:

- designed to collect customer and competitive intelligence information;
- undercover or anonymous nature of the job;
- discloses the process of service provider–customer interaction rather than the outcome; and
- represents a learning opportunity for an organization, as the MC technique is rooted in third-party feedback, and so it is logically the element most closely related to an actual client.

Apart from the quality and quantity of the collected and reported data by the MC agent, there are two relevant criteria to take into consideration in the design and planning of the MC process: reliability and credibility. If subsequent measurement repetition of the same observation produces a similar outcome under similar conditions, we may classify that measurement as reliable (Trochim and Donnelly, 2007). That means that the obtained result is trustworthy, thus it is also credible. Such credibility is based on the evidence of objectivity of the measurement process in opposition to the subjectivity (prone to error) of the agent who performs it (Bocking, 2004).

Received 8 January 2015
Revised 29 April 2015
20 August 2015
Accepted 7 October 2015

The current issue and full text archive of this journal is available on Emerald Insight at: www.emeraldinsight.com/0887-6045.htm
Mystery client traits on service evaluation

Table I Mystery client definitions

<table>
<thead>
<tr>
<th>Authors</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hurley (1998)</td>
<td>“Mystery shopping is an approach that aims at getting service providers to see their services as consumers see them… anonymous professionals show up as customers”. p. 594</td>
</tr>
<tr>
<td>Ford et al. (2011)</td>
<td>“Mystery shoppers go to an organization to experience a service ‘incognito’. They act as a typical customer to assess predetermined service standards of service quality are being met by employees delivering the service”. p. 159</td>
</tr>
<tr>
<td>Cook et al. (2002)</td>
<td>“Mystery shopper is a surrogate guest/customer who, unbeknownst to the service provider. Evaluates the service against ‘ideal’ and then objectively reports out on their experience”. p. 170</td>
</tr>
<tr>
<td>Butcher et al. (2009)</td>
<td>“Mystery shopping involves an outsider to visit the accommodation as a guest and report on how they perceived the level of service on key attributes”. p. 392</td>
</tr>
<tr>
<td>Tarantola et al. (2012)</td>
<td>“They act as normal or potential customers and make unannounced visits to the company… to evaluate and monitor customer satisfaction and quality of service in different sectors”. p. 10104</td>
</tr>
<tr>
<td>León et al. (1994)</td>
<td>“Mystery-client technique method is characterized by a new ‘client’, who controls the client- provider interaction as she or he solicits services and is not obtrusive, since the provider is unaware that the client is conducting a test”. p. 185</td>
</tr>
<tr>
<td>Finn and Kayandé (1999)</td>
<td>“Typically, the observer enters the outlet to be evaluated posing as an average customer and, immediately after engaging in what appears to be a normal customer interaction, completes a detailed report on various aspects of the service and shopping environment at the outlet”. p. 1</td>
</tr>
<tr>
<td>Wiele et al. (2005)</td>
<td>“Organizations can, for example, measure the quality of service delivery by making use of mystery guests, which are well-trained people who behave as normal customers but who are accurately observing what is going well and what can be improved in the service process as perceived by them”. p. 533</td>
</tr>
<tr>
<td>Wilson (1998)</td>
<td>“Mystery shopping, a form of participant observation, uses researchers to act as customers or potential customers to monitor the processes and procedures used in the delivery of a service”. p. 148</td>
</tr>
</tbody>
</table>

Ultimately, under the service quality narrative and framework (Ranjian et al., 2015; Urban, 2013), MC helps the organizations to control operation management efficiency, to detect problems and to avoid service failures. That specific topic will be examined in the next section.

1.2 Justification of mystery client method

The rationale for service evaluation activity is quite straightforward: quality and innovation. Service quality and the implementation of an innovation culture are the recognized strategies for improving competitive advantage, as well as the key area customers focus on during the service encounters (Bowen, 1990; Bitter, 1990; Jayawardhena, 2010; Kim et al., 2012). Service evaluation is neither static nor universal, but appraised on a wide range of factors varying by provider-to-provider and environment-to-environment. However, although evaluations are subjective, the vast numbers of service evaluation frameworks provide comprehensive and thorough evaluation frameworks (Seth et al., 2005). Attributes evaluated range from tangible physical elements such as service encounter appearance, waiting time and product display, to abstract intangible elements like perception of employee responsiveness, competence and satisfaction. Grönroos (1984) described services on two dimensions: functional (what is provided) and technical (how it is provided). Specifically, functional evaluation is the outcome related to whether customers receive utility from service interactions. Rust and Oliver (1994) extended this conceptualization to include environmental factors (conditions under which the service is performed). Farmer (1988) proposed a Service Attributes Quality Model, which bases service quality on three basic attributes: physical facilities and processes, people behavior and customer professional judgment. McDougall and Levesque (1994) added a fourth dimension to service evaluation, which involves evaluating encounters based on the ease of accessing the service experience (enabling). Brady and Cronin (2001) introduced the concept of interpersonal and administrative interaction into service experience evaluation. Dagger et al. (2007) synthesized a hierarchical service evaluation model specific to the health-care industry and raises concerns of transferability to other service areas.

A modified version of Parasuraman et al.’s (1994) model recognized the multi-dimensionality of service evaluation, and future developments adopted an integrative perspective by analyzing service performance (Brady and Cronin, 2001; Dagger et al., 2007) and incorporating objective factors such as service quality environment and quality of processes (Zhang et al., 2012; Urban, 2013).

1.3 Research goals

The quest to determine to what extent this method is objective, an accurate or reliable measure of the experience under observation has been attempted by several researchers. Service encounter is the strategic moment when because of an interaction process, the service is provided and the transaction is accomplished. The actor who experiences that service in terms of flow of time and performance judgment is the client (Cook et al., 2002). Consequently, the role of (mystery) client as a collecting data entity is necessarily relevant. Morrison et al. (1997) discussed social and cognitive aspects, mostly relying on the different memory mechanisms of encoding, storage and retrieval of observed target information. Wilson (1998, 2001) analyzed the factors/techniques used to maximize the reliability of this diagnostic tool and the managers’ justification for using it. Calvert (2005) criticized the technique, as it distorts reality because it considers a snapshot of service in a specific place, at a specific time. Other
authors pinpointed that individual demographical and behavioral characteristics can influence the quality of data collected (Wilson, 1998; Erstad, 1998; Finn and Kayandé 1999; Dagger et al., 2007). León et al., (1994) went further in proving an answer to the quest. Their investigation was the closest to our research goals. They trained 18 women to work in pairs as simulated clients to observe and rate 84 family planning specialists. By measuring the inter-rater agreement of six relevant variables, they found that the checklist-item scores were better than ratings. Although, methodologically insightful, this research was associated with health-care activities thus unrelated to service marketing context.

The aim of the study is to examine the impact of the effect of personal characteristics on the accuracy and consistency of MC technique. Let us consider the following scenario: two observers are simultaneously experiencing a service; if they belong to the same market segment and share similar demographic and psychographic characteristics, they should, in principle, evaluate that service in an exactly identical way. That is the main assumption for the sake of the scientific and managerial validity of the MC method. An MC is not a “true” client but someone who pretends to be and performs the role. Theoretically, to understand the rationale behind the ability to perform an MC task, we rely on two major explanatory branches: multiple personas and social role. Those two theoretical domains are discussed in the next section.

2. Theoretical foundations

2.1 Multiple personas
Some degree of variability in the way individuals express their personality seems to be a natural human trait. Excluding schizophrenia or psychotropic multiple personality disorder symptoms, multiple personality behavior coexists within a well-adjusted and coherent ego identity or core self (Raggatt, 2000). The self-theory discussion has helped to explain the integration and sane conviviality among those personas, but the self-construction process is also a function of significant others’ representations (Andersen and Chen, 2002). The diversity of personas is dependent on interpersonal contextual cues to induce a repertoire of self-regulated relational selves, which is particularly relevant and manifest in consumer behavior (Bahl and Milne, 2010; Gould, 2010). The process of generating multiple personas is dynamic. If we rely on our past experience self-schemas to build those personas, we also have to nurture and scaffold possible selves to cope with future situations (Markus and Nurius, 1986).

Accepting the evidence that human beings exhibit such personality plasticity, the number of potential personas an individual can develop and reveal may vary. The self-concept differentiation was defined by Donahue et al. (1993, p. 834) as “the tendency to see oneself as having different personality characteristics in different roles and psychological adjustments”. Holding a high self-concept differentiation means having a highly fragmented identity to flexibly and adaptively respond to distinct role requirements. That division of self implies emotional and social relationship adjustments, and is more common in the contexts of frequent social role change and role performance dissatisfaction.

A possible psychological explanation of multiple personas is the simultaneous management of multiple impressions (Leary and Allen, 2011). A process designated as self-presentation persona reflects a person’s desire to project specific images deemed appropriate to a particular target or situation. The number of self-presentation persona can range from a distinct mode, conveying a more idiosyncratic spin in tune with the individual’s self-views to a normative mode. Within the latter, the size of repertoire of personas varies according to the degree of agreeableness, self-esteem and authenticity or Machiavellianism stance.

2.2 Role theory
Because the multifaceted nature of people explains the reason for multiple personas, the social instrumentalization of this characteristic has been studied by a social role theory. Rooted in the division of labor concept and based on the dramaturgical metaphor, the social role determines distinct occupational positions for the agents corresponding to a specific job status, thus the person and the role are independent concepts (Biddle, 1986).

The role represents a social construction which is structurally defined to fulfill not only biological needs but also work-related requirements. Apart from the stereotyped sex role classification, both genders can simultaneously occupy the roles of parents, care provider, mother/father, wife/husband and daughter/son (Barnett et al., 1992; Martire et al., 2000). In addition to the clear evidence of role accumulation in the non-normative condition, adult individuals execute work-related activities as employees in a hierarchical organization, being simultaneously subordinate vis-à-vis her/his supervisor and leaders of a team (Hong and Seltzer, 1995).

Occupying multiple roles demands resources of time and energy, and regardless of the flexibility span, those assets are limited, thus requiring some degree of work–family tradeoff (Greenhaus and Beutell, 1985). A common consequence of playing multiple roles within an organization is role conflict and role ambiguity occurring because of an expectation/knowledge deficiency about each manager-specific role (MaGee et al., 1989; Shepherd and Fine, 1994). Drawing on the literature of consumer behavior/decision-making, buyer experience in a service encounter also represents a well-defined role where both service employees and consumers participate like actors on stage, performing according to a “service script” (Guiry, 1992) and facing similar psychological and sociological constraints (Otnes et al., 1997). Buyers and employees behave interdependently and complementarily. Each should understand the other’s positions and predicted role, as role expectations affect service quality and customer satisfaction (Solomon et al., 1985).

If a human being inherently assumes different personas following specific situational cues or responding to her/his intrinsic choices and is also able to play multiple roles, then how can she/he compartmentalize such diversity among personas/roles that are not always congruent? To what extent is she/he competent to psychologically adjust or reorganize her/his personality characteristics according to the most appropriate role? This dispute around the person–role merger magnitude (Turner, 1978) raises a question about consistency (identity theory – Jackson and Smith, 1999) versus appearance (role theory – Biddle, 1986) in the process of “wearing in” and
“taking off” different roles. Surely, to attain the MC method objectives, the further and deeper the merger process is potentially more accurate will be her/his evaluations.

### 3. Mystery client potential individual differences

Behavioral dimensions of service operations should not be undermined, as regardless of how mechanistically employees are programmed to manage their activities, there are always in some extent personal contact and interaction with someone (Croson et al., 2013). Both (mystery) clients and employees express and react emotionally which may influence their response/behavior (Van Eerde and Peper, 2008; Wang et al., 2012; Urda and Loch, 2013).

Previous research on consumer behavior literature extensively examines the influence of individual characteristics on service evaluation (Iacobucci and Ostrom, 1993; Hopkins et al., 2009). Some relevant characteristics emerging are demographics and psychographic. Given the interactive and personal nature in production and consumption of services, these traits are considered especially relevant to service providers. We adopt a conceptualization of individual traits to emphasize distinctive qualities of an individual which shape personal character. Based on a review of the literature, personal characteristics are expected to have a significant influence on service evaluation (Dabholkar and Bagozzi, 2002). Six characteristics were examined (five psychological and one demographical) to ascertain the statistical invariance in the MC rating. Each of these factors is relevant in consumer-related research from previous empirical studies. The hypotheses rooted on the theoretical predictions are formulated in next sub-sections.

#### 3.1 Gender

There is consensus that men and women act differently in their environment and evaluate their encounters differently (Iacobucci and Ostrom, 1993). Men are more likely to pay attention to functional components of their service experiences, while women pay more attention to relational aspects of their encounters. Therefore, in rating services, men are likely to pay attention to attributes that pertain to the core service, while relationship attributes are more salient to women (Sharma et al., 2012). Although, to the best of our knowledge, the impact of gender difference on MC has not been examined, the literature suggests that gender might have some influence on personal evaluation (Levy, 1988):

- **H1.** The association between MC’s gender and each service evaluation variable is statistically significant.

#### 3.2 Opinion leadership

Reynolds and Darden (1971) suggested that individual difference relates to a person’s inherent leadership characteristic. Generally, opinion leaders have a dominating influence on the behavior of others and in particular, members of their social group. They are translators of marketing stimuli and act as intermediaries between the marketer and social peers. Opinion leaders are self-confident and share product information with members of her/his social peer. These characteristics manifest in the shopping practices. For instance, customers with more product-related knowledge are more likely to have enhanced information-processing capabilities (Alba and Hutchinson, 1987; King and Balasubramanian, 1994; Bont and Schoormans, 1995). The influence of opinion leadership trait on MC evaluation remains, to the best of our knowledge, largely unanswered. MCs who are opinion leaders will have a clear idea of what to expect from their service experience encounters and will be more assertive in their evaluations taking into account more details. It is in this regard that MC will be more vigilant and less tolerant for perceived non-conformities:

- **H2:** 1 The ability to influence others (component of opinion leadership) is positively associated with the ratings of each service evaluation variable.
- **H2:** 2 The agreeableness with the information received from others (component of opinion leadership) is not statistically associated with the ratings of each service evaluation variable.

#### 3.3 Consumer involvement

Consumer involvement towards a product category reflects the degree of personal relevance and importance (Zaichkowsky, 1994). However, the conceptualization of product involvement has evolved in recent times and extended to include affective and emotional components (Zaichkowsky, 1994; Kim and Sung, 2009). Relative to the impact of involvement in service evaluation, Goodman et al. (1995) concluded that customers who are more involved with a product/service category are more likely to see each aspect of the service encounter as personally relevant. It is reasonable to expect therefore that MC, who consider service category as highly relevant to their life, would be more inclined to evaluate service attributes in detail:

- **H3:** 1 The involvement in product category is positively associated with the ratings of each service evaluation variable.
- **H3:** 2 The product ignorance (component of consumer involvement) is not statistically associated with the ratings of each service evaluation variable.

#### 3.4 Need for emotions

The need for emotion describes an individual propensity to actively seek out and enjoy emotional stimuli, and have an expressed preference to use their own emotions during personal interactions (Raman et al., 1995). Generally, consumers with higher emotional propensity exhibit emotional responses to environmental stimuli (Harris and Moore 1990) and tend to evaluate service encounters more positively compared to individuals who experience negative emotions or are neutral. Gardner (1985) contends that a positive mood increases the likelihood that performance and behavior are evaluated on the positive end, whilst a negative mood is more likely to influence a negative evaluation of service outcomes:

- **H4:** The association between need for emotion and each service evaluation variable is statistically significant.
3.5 Faith in intuition
Faith in intuition was developed by Epstein et al. (1996), originally named intuitive–experiential scale, and it correlates significantly with personality traits of conscientiousness and esoteric thinking. According to Epstein et al. (1996), there are two systems of human processing, namely, rational thought and experiential thinking, which influence each other. The rational system is a conscious, analytical, slower and relatively free of emotion. In contrast, the experiential system is a pre-conscious, fast, automatic, holistic and associated with affection. Following this concept, we refer to faith in intuition as an intuitive spirit that measures the confidence in one’s feeling and immediate impressions as a basis for decision-making:

H5. The association between the components of faith in intuition and each service evaluation variable is statistically significant.

3.6 Need for cognition
Cacioppo and Petty (1982) described a persons need for cognition as the tendency to engage in and enjoy cognitive deliberations. The authors explained that individuals with higher cognition preferences tend to scrutinize marketing stimuli, engage in extensive elaboration and reasoning and enjoy relatively effortful cognitive tasks. These individuals pay attention to intricate details on messages and environmental stimuli when forming opinions and attitudes (Cacioppo et al., 1982). These individual differences can influence how people understand, evaluate and interpret their environment and affect every facets of marketing. On the other hand, individuals with lower cognitive need are more passive in their approach and lack the ability and motivation to think extensively about the stimuli:

H6. The association between the components of need for cognition and each service evaluation variable is statistically significant.

3.7 Overview
We have two streams of theoretical elaborations to interpret the prospect of service evaluation differences between MC:

1. **Rationalist view:** As well-trained professionals MC tend to strictly follow the prescribed procedures by mechanically repeating previously learned role script and pursuing an optimal decision-making goal to guide their activities (Yanof, 2012; Croson et al., 2013). Because the MC job design is highly standardized with predefined evaluation criteria and structured responses, we expect that the MC evaluation will always be equal regardless of individual demographical and psychological profile.

2. **Behavioral operations view:** MC professionals are also human beings; although acting in a deliberate manner, there are necessarily individual differences which hardly lead to uniform results or to a unique outcome.

Even when environment and contextual factors were under control according to pre-defined parameters, personal heterogeneity among (MC) agents produces some variability (Croson et al., 2013; Urda and Loch, 2013).

To what extent do the service evaluations’ ratings vary as a function of the differing MC’s psychographical and demographical profile? Assuming that all MC participants in our study perform the same task, share the same training, role script and data-collecting instrument and motivation, are there statistical significant differences in their ratings between two mutually exclusive MCs’ profile? Which theoretical view is more likely to pertain? According to the rationalist view, the hypothetical answer would be “no”. If the behavioral operations view is retained, we expect that a variety of ratings will be in tune with MCs’ heterogeneity.

4. Methodology

4.1 Research design and statistical techniques
The data collection mirrored the sequential approach of Erstad (1998). The MCs were recruited from a sales management course offered to students pursuing master’s qualifications at a local university. The use of students is justified in this case, as they represent a homogeneous group in terms of education and age. The potential MCs received intensive training designed to provide skills and expertise on how to conduct effective MC surveys. At the end of the training session, each student was assigned to a group that consisted of three to four members. Further, 36 retail stores were selected for evaluation based on two selection criteria: type of product sold (retailers of consumer electronic products) and location of the retail store (located in Porto Metropolitan area, Portugal). Overall, 144 MCs conducted 355 evaluations of the stores and salespersons. Over 200 h of contact was spent between mystery shoppers and salespersons, and this facilitated cross comparisons, given that the MCs rated different stores. Given the independent multiple visits, the comparisons were not biased. Hair et al. (2010) advised that sample sizes should achieve a minimum of five observations for each dependent variable; in our case, given that we evaluated eight personal characteristics (seven psychological and one demographical), the minimum sample criterion was satisfied. To maintain consistency and control, each group interacted with the same salesperson in each store. The MC adopted the role of prospective buyers of electronic products (laptop, MP3/4 player or digital camera). Each member played a specific role: one acted as the prospective buyer, while the others played the role of accompanying friends and had the task of observing and recording data, to reduce the memory burden on the MC (Morrison et al., 1997). Relative to the data-collection process, the type of interaction between the MC and service personnel was standardized and scripted. This allowed consistency with interaction but also gave the salesperson the opportunity to act freely as possible. The same MC team visited at least two other stores, allowing each member to perform different tasks active versus passive observation/interaction. Simultaneously, each of the 36 stores was visited by at least two teams. That is the same employee was observed several times. Relative to interaction, the MCs were required to wait for 5 min maximum near the location of the product they were interested in purchasing. If the MC received no assistance within 5 min, the MC then actively sought help. All participants memorized a script assuring that they verbalized exactly the same content in the same way – non-verbal.
language included (Appendix). Immediately after the service encounter, each MC independently completed a standardized service evaluation form that evaluated eight specific criteria rooted in Parasuraman et al.'s (1994) service quality model and satisfaction with the overall service encounter: salesperson empathy, assurance, responsiveness, reliability, appearance of people, appearance of physical, environment satisfaction and service adhering to expectations. The service evaluation instrument consisted of 32 items measuring these 8 facets of the MC experience on a seven-point Likert scale. To maintain consistency, negatively worded items were reversed coded. The first six were measured on multi-item scales and the last two on a single-item scale. The use of Parasuraman et al.'s (1994) service quality model is not new to MC research and has been used in multiple service evaluation studies (Dawes et al. 2000; Wiele et al., 2005).

To analyze the data, the study used a series of steps to condense, purify and categorize the data. This process started with analyzing the MC personality using exploratory factor analysis (EFA) and k-mean non-hierarchical cluster analysis. Second, the analysis of variance (ANOVA) was used to determine whether group differences existed on each of the 32 evaluation criteria. This facilitated cross-referencing of the average stores and facilitated the detection of group differences. For example, Store 2 (including four salespersons working there) was evaluated ten times by three separate MC groups. Each group had, on average, 3.4 members. The rating of each group was calculated and cross-referenced to other groups who visited the same store. This referencing was carried out on each service variable measured. The cross-referencing was set at 95 per cent confidence level and sought to determine whether variations existed in store evaluation. Multiple regression analysis was subsequently used to ascertain the significance of each characteristic in explaining MC evaluation, and multiple analysis of variance (MANOVA) was used to examine the statistical difference between varying MC profiles. As the rationale behind this research is to assess if the individual traits affect the way they evaluate service providers, a necessary condition to assure accuracy in that measurement is to control the sources of variance. We accomplished that goal in three ways:

1. the use of a homogeneous group of MCs in terms of education and age;
2. at least one-third of the observations/evaluations were made in the same stores and with the same salesperson by several teams of MCs ensuring that all were submitted to the same stimuli; and
3. the time span of that observation/evaluation was one week which rendered unlikely variations in store environment and changes in retail or product/category promotional activities.

There are other methodological alternatives. An experimental research design under laboratory settings would guarantee that all participants watched the same video exhibiting in-store environment and a sales process interaction simulation. Still, the lack of ecological validity would undermine that option. If all MC participants visited a single store and interacted exactly with the same salesperson, it would fail effectiveness because of the suspicion about the intention. The mystery circumstance quickly ceased to work. To overcome such limitations, we considered 36 stores (real context) and multiple visits of different MC teams interacting with the same salesperson allowing to generate data and making it feasible to apply to the multivariate analysis. Therefore, underlying the service/store evaluation (as dependent variables) and MCs’ profile (as independent variables) relationship, the unique source of variance was the MCs’ individual traits.

4.2 The measurement instrument

4.2.1 Individual profile

The administration of the profiling instrument occurred at the briefing session. The design of the instrument facilitated the subsequent grouping of MCs into distinct psychographic profiles. The instrument consisted of 82 items, which were MCs’ disposition on eight characteristics (seven psychological and one demographical), namely: opinion leadership (with two components), level of involvement in purchase decision (with two components), need for emotions, faith in intuition and need for cognition and gender. For measurement, the study used the constructs adopted from previous studies (Table II).

4.2.2 Service evaluation

The service evaluation instrument consisted of 32 items measuring 8 facets of the MC experience on a seven-point Likert scale. The service attributes measured aspects of service quality dimensions proposed by Parasuraman et al. (1988) but modified it to fit the context of the study. Specifically, the instrument evaluated salesperson empathy, assurance, responsiveness, reliability, appearance of people, appearance of physical environment, satisfaction and service adhering to expectations (Appendix). To maintain consistency, negatively worded items were reverse-coded. The first six were measured on multi-item scales and the last two on a single-item scale.

5. Findings

In total, 52 per cent were female. On average, MC participants were 23.14 years old (SD = 2.99).

5.1 Determining the mystery client profile

First, EFA using the principal component approach with varimax rotation was used as the primary data-reduction technique. The use of EFA is justified given its ability to systematically reduce a number of variables to simplify the analysis and maintain parsimony. Items with low correlations \( r < 0.30 \) and communalities \( <0.50 \) were removed from further analysis. To determine the reliability, Cronbach’s alpha statistic was calculated for the five psychographic characteristics. The first characteristic was Opinion Leadership on eight items adopted from Reynolds and Darden (1971). One item was deleted from further analysis because of low correlations. The retained items generated a two-factor solution and explained 78 per cent of total variance. Factor 1 comprised four variables relating to participant’s ability to influence others. The reliability for this retained scale was moderately high – 0.934. For each of the subsequent constructs, Table III summarizes the percentage of total variance explained (EFA output), their components, reliability value (Cronbach’s alpha) and the corresponding items. Comparing with the original structure (Table III), some of the
items were dropped because of the low statistical significance ($p > 0.1$) and inter-item correlations.

The second stage involved classifying MCs into homogeneous and distinctive groups based on the personality traits derived from the first stage of our analysis (factor analysis described above). To define the classifications, the study used a non-hierarchical k-mean clustering approach. The k-mean approach is a common method to partition data sets into predefined number of groups or k groups (Wagstaff et al., 2001). The process involved three steps: first, we defined the amount of groups (K) and using a random process to generate cluster centers for each group. Second, each observation was assigned to one of the clusters based on similarity of each center and iteratively refined, so that each observation resides closest to its respective center. Finally, each center is updated to reflect the mean of the observations classified into the specific group. The objective of this approach revolves on grouping MCs into taxonomies based on personal characteristics. Two groups (K = 2) were defined for each personal characteristic (Group 1 – high and Group 2 – low). The numbers indicate how strong the MC ratings were on that particular characteristic. MCs with stronger ratings were classified into Group 1 – or high category, whereas lower ratings were classified into Group 2 – low category. A standard $F$-test showed that the difference between both groups was statistically different on each personal characteristic. Table IV summarizes the mean score of the high and low groups for each personal characteristic.

### 5.2 Analysis of service evaluation criteria by mystery client group: cross analysis
The general aim of the cross-referencing analysis was to explore whether service evaluation criteria can vary across individual groups. Given that our MC evaluated different stores, it was imperative to test the consistency of MC service ratings. The study cross analyzed the ratings of each MC on each service criteria using one-way ANOVA analysis. The purpose of the ANOVA was to examine whether the MC ratings were related to individual characteristics. ANOVA compared the mean rating of each store on the 32 items on the questionnaire which measured 8 service dimensions outlined in our research model (assurance, empathy, tangible-place, reliability, responsiveness, conformance to expectations and satisfaction). The research design ensured that each group visited at least two stores previously visited by other groups and interviewed the same salesperson, hence conceptually justifying the cross analysis. In addition, the research was designed with a relatively balanced group size which allowed us to identify significant differences in MC ratings. In terms of process, the cross analysis involved the following stages:

- For each store visited, each group’s average rating on each service item was compared.

#### Table II Individual characteristics

<table>
<thead>
<tr>
<th>Type</th>
<th>No. of items</th>
<th>Measurement scale</th>
<th>Reported reliability</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opinion leadership</td>
<td>8</td>
<td>Seven-point Likert (1 – Strongly disagree to 7 – Strongly agree)</td>
<td>0.79 and 0.73</td>
<td>Opinion leadership is reflected in two basic dimensions: first, the extent to which a person exerts influence and gives information to others on his/her own opinions and second, the extent to which the person actively seeks information from others</td>
</tr>
<tr>
<td>Consumer involvement profile (CIP) (Laurent and Kapferer, 1985)</td>
<td>16</td>
<td>Five-point Likert (1 – Strongly disagree to 5 – Strongly agree)</td>
<td>0.80; 0.90; 0.88; 0.82 and 0.72</td>
<td>Conceptualize involvement as a multifaceted construct determined by five antecedents, namely: perceived importance of the relevance of the product; perception of risk in purchase which relates to the consequences of mis-purchase; the symbolic value assigned to the product class; hedonic or emotional value attributed to the product class; and interest in an enduring relationship with the product class.</td>
</tr>
<tr>
<td>Need for emotion (NEF) (Raman et al., 1995)</td>
<td>12</td>
<td>Five-point Likert (1 – Strongly disagree to 5 – Strongly agree)</td>
<td>0.87</td>
<td>Evaluates individual propensity to look for emotional situations and stimuli and to use emotion in their interaction with world. The scale focuses on short-term emotions.</td>
</tr>
<tr>
<td>Faith in Intuition (FI) (Epstein et al., 1996)</td>
<td>12</td>
<td>Five-point scale (1 – Completely false to 5 – Completely true)</td>
<td>0.87</td>
<td>It measures confidence in decision making which is reflected in several aspects including personality, adaptation, self-fulfillment, interpersonal relations.</td>
</tr>
<tr>
<td>Need for Cognition (NFC) (Cacioppo and Petty, 1982)</td>
<td>34</td>
<td>Nine-point Likert (4 – Completely disagree to 4 – Completely agree)</td>
<td>0.80</td>
<td>Relates to the extent to which an individual engages in and enjoys thinking.</td>
</tr>
</tbody>
</table>
Table III  Measurement of individual characteristics

<table>
<thead>
<tr>
<th>Construct (%total expl.)</th>
<th>Components (Reliability)</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opinion leadership (78%)</td>
<td>Ability to influence (0.934)</td>
<td>My friends and neighbors often ask my advice about electronic/digital products I sometimes influence the type of electronic/digital products my friends buy I feel that I am generally regarded by my friends and neighbors as a good source of advice about electronic/digital products My friends usually come to me, more than I go to them, for information about electronic/digital products Agreeableness (0.770) I often seek out the advice from my friends regarding which electronic/digital products I should buy I spend a long time talking to my friends regarding which electronic/digital products I should buy My friends and neighbors usually give me good advice on what brands of electronic/digital products I should buy</td>
</tr>
<tr>
<td>Consumer involvement profile (71%)</td>
<td>Involvement (0.874)</td>
<td>Buying computers/related products is like buying gifts for myself It gives me pleasure to purchase electronic/digital products I attach a great deal of importance to electronic/digital products One can say electronic/digital products interest me a lot Ignorance (0.787) Choosing electronic/digital products is rather complicated When I face a shelf of electronic/digital products, I always feel a bit at a loss to make my choice When one purchases electronic/digital products, one is never certain of one’s choice</td>
</tr>
<tr>
<td>Need for emotion (70%)</td>
<td>Need for emotions (0.704)</td>
<td>Experiencing strong emotions is not something I enjoy very much I look forward to situations that are less emotionally involving I don’t look forward to being in situations that others have found to be emotional I prefer to not get involved with the emotional aspects of any situation Need for cognition (72%) Cognition preference (0.703) I would rather do something that requires little thought than something that challenges my thinking abilities I find little satisfaction in deep and lengthy deliberations Thinking is not my idea of fun Cognition practice (0.707) I like tasks that require little thought once I’ve learnt them I only think as hard as I have to Need for cognition (72%) Cognition practice (0.707) I like tasks that require little thought once I’ve learnt them I only think as hard as I have to</td>
</tr>
<tr>
<td>Faith in intuition (66%)</td>
<td>Faith in others (0.8115)</td>
<td>My initial impressions of people are almost right I trust my initial feelings about people I am a very intuitive person I believe trusting to my hunches When it comes to trusting people, I can usually rely on my “gut feelings” Need for cognition (72%) Cognition preference (0.703) I would rather do something that requires little thought than something that challenges my thinking abilities I find little satisfaction in deep and lengthy deliberations Thinking is not my idea of fun</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Self assured (0.790) I often have a clear visual image of things I am good at visualizing things Need for cognition (72%) Cognition preference (0.703) I would rather do something that requires little thought than something that challenges my thinking abilities I find little satisfaction in deep and lengthy deliberations Thinking is not my idea of fun</td>
</tr>
</tbody>
</table>

Table IV  Non-hierarchical cluster analysis

<table>
<thead>
<tr>
<th>Personal characteristics</th>
<th>Group 1: Group 2:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Char. 1 Ability to influence (Op.Lead)</td>
<td>0.519** -1.477**</td>
</tr>
<tr>
<td>Char. 2 Agreeab (Op.Lead.)</td>
<td>0.570* -1.269*</td>
</tr>
<tr>
<td>Char. 3 Involv (Cons.Inv.)</td>
<td>0.728** -0.956**</td>
</tr>
<tr>
<td>Char. 4 Ignorance (Cons.Inv.)</td>
<td>0.791** -0.88**</td>
</tr>
<tr>
<td>Char. 5 Need for emot (NFE)</td>
<td>1.13** -0.595**</td>
</tr>
<tr>
<td>Char. 6 Faith in others (Fi)</td>
<td>0.673** -0.743**</td>
</tr>
<tr>
<td>Char. 7 Self assured (Fi)</td>
<td>0.456** -1.26**</td>
</tr>
<tr>
<td>Char. 8 Cogn Prefe (NFC)</td>
<td>0.749** -0.862**</td>
</tr>
<tr>
<td>Char. 9 Cogn practice (NFC)</td>
<td>1.05** -0.645**</td>
</tr>
<tr>
<td>Char. 10 Trust in Cogn (NFC)</td>
<td>0.934** -0.745**</td>
</tr>
</tbody>
</table>

Notes: Mean contrast are significant ( **p < 0.05; *p < 0.1) according to the F-test

For each store visited, an ANOVA compared the average rating of each group on each service item at a significance level of 95 per cent. ANOVA allowed us to identify statistical differences between our group means by using the F-ratio.

In summary, the result showed that there are significant differences in intergroup comparison on each service evaluation criteria. This means that individuals who rated the same store, varied in their service evaluation. By statistically comparing the different MC groups (p < 0.05), Table V shows the 32 service evaluation attributes. The most profound influence occurred in the perception of physical store appearance (dusty) and the lowest perception of salesperson ethics. This provided the evidence that suggested significant intra-personal influence on service evaluation that may be unique to an individual, which suggests the need to engage in
further analysis using multivariate techniques such as multiple regressions and MANOVA.

5.3 Creating summated scales for service evaluation criteria

Although the use of multi-item scales enhance precision and improve reliability, multiple items lends complication to data analysis (Spector, 1992). In this regard, to maintain parsimony and simplify the analysis, summated scales were constructed. Summated scales involve collapsing the data into a single variable. In our study, the service evaluation instrument consisted of 32 questions representing the eight attributes of service quality we intended to measure. In this regard, we simplified the MANOVA and regression analysis by collapsing the 32 items into 8 summated scores. Summated scales were calculated by taking the average of the items in each scale. To account for internal consistency, unidimensionality of each scale was verified through EFA and the Cronbach’s alpha statistic was calculated. Furthermore, to adequately compare the relative effects of personality characteristics on each service attribute, we standardized the ratings. Standardization converts the variables into a common scale (mean = 0; SD = 1) making the variables comparable (Hair et al., 2010).

Table VI presents the descriptive statistics on each service attribute, including the standardized mean and standard deviation and reliability index.

5.4 Multiple regression analysis

To analyze the relationship between each service evaluation criterion and MC profile, the study used multiple regression analysis. The technique involved assessing the character of the relationship between each of service attribute (dependent variable) and personal characteristics. To maintain parsimony, confirmatory perspective to multiple regressions was used where the individual characteristics were specified as independent variables. Eight regression variates were formed (one for each service attribute). The regression variates are represented below:
Table VI  Service evaluation criteria

<table>
<thead>
<tr>
<th>Store evaluation criteria</th>
<th>No. of items in the service attribute</th>
<th>Summated score Mean</th>
<th>SD</th>
<th>Chronbach’s alpha Mean</th>
<th>Standardize score Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extent to which service encounter conformed to expectations</td>
<td>1</td>
<td>3.18</td>
<td>0.946</td>
<td>–</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Satisfaction with service encounters</td>
<td>1</td>
<td>3.37</td>
<td>0.995</td>
<td>–</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Empathy</td>
<td>10</td>
<td>42.67</td>
<td>7.83</td>
<td>0.845</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Assurance</td>
<td>5</td>
<td>23.73</td>
<td>6.4</td>
<td>0.887</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>4</td>
<td>31.15</td>
<td>5.67</td>
<td>0.690</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Reliability</td>
<td>3</td>
<td>14.68</td>
<td>3.62</td>
<td>0.752</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Tangible evidence – place</td>
<td>6</td>
<td>30.61</td>
<td>6.40</td>
<td>0.738</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Tangible evidence – people</td>
<td>2</td>
<td>10.74</td>
<td>2.80</td>
<td>0.604</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: Number of stores visited = 36 (p < 0.05)

\[
Y_i = ABI + AG + INv + IG + NFE + FIO + SA + CP + CPR + TIC + \text{Gender}_{0, \text{Male}, 1, \text{Female}}
\]

Where:

- \( Y_i \) = Ratings on each service attribute (salesperson empathy, assurance, responsiveness, reliability, appearance of people, tangibles, satisfaction and service adhering to expectations);
- \( ABI \) = Ability to Influence (Opin. Leader);
- \( AG \) = Agreeableness (Opin. Lead);
- \( INv \) = Involvement (Cons. Invol.);
- \( IG \) = Ignorance (Cons. Invol.);
- \( NFE \) = Need for emotions;
- \( FIO \) = Faith in others (FI);
- \( SA \) = Self-Assured (FI);
- \( CP \) = Cognitive preference (NFC);
- \( CPR \) = Cognitive Practice (NFC);
- \( TIC \) = Trust in cognition (NFC);
- \( \text{Gender} \) = (Male; Female).

The results of the regression analysis are summarized in Table VII.

Table VIII shows that significant relationships (p < 0.05) between MC involvement in product category and the ratings assigned to four service evaluation criteria: empathy, assurance, responsiveness and perception of service personnel. This suggests that MCs with different levels of involvement assigned different ratings on three service evaluation criteria.

To ascertain the nature of this variation, we used MANOVA in the two profiles (high and low involvement) confirming that MCs’ involvement in the product category was a significant discriminator in service evaluations (Roy’s largest root \( p = 0.002 \)). MCs were more likely to assign higher ratings on the

Table VII  Multiple regression analysis

<table>
<thead>
<tr>
<th>( Y )</th>
<th>Extent to which service encounter conformed to expectations</th>
<th>Satisfaction with service encounters</th>
<th>Empathy</th>
<th>Assurance</th>
<th>Responsiveness</th>
<th>Reliability</th>
<th>Tangible evidence – place</th>
<th>Tangible evidence – people</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to influence (Op. Lead)</td>
<td>0.017</td>
<td>0.078</td>
<td>-0.028</td>
<td>0.049</td>
<td>-0.012</td>
<td>0.064</td>
<td>0.107</td>
<td>-0.027</td>
</tr>
<tr>
<td>Agreeab (Op. Lead.)</td>
<td>-0.058</td>
<td>-0.032</td>
<td>-0.006</td>
<td>0.050</td>
<td>-0.041</td>
<td>-0.009</td>
<td>0.069</td>
<td>0.027</td>
</tr>
<tr>
<td>Involv (Cons. Inv.)</td>
<td>-0.110</td>
<td>-0.061</td>
<td>-0.084</td>
<td>-0.125</td>
<td>-0.141**</td>
<td>-0.036</td>
<td>0.095</td>
<td>-0.118</td>
</tr>
<tr>
<td>Ignorance (Cons. Inv.)</td>
<td>-0.017</td>
<td>-0.029</td>
<td>-0.035</td>
<td>0.050</td>
<td>-0.064</td>
<td>0.041</td>
<td>0.016</td>
<td>0.042</td>
</tr>
<tr>
<td>Need for emot (NFE)</td>
<td>0.018</td>
<td>0.023</td>
<td>0.000</td>
<td>-0.045</td>
<td>-0.037</td>
<td>0.021</td>
<td>-0.001</td>
<td>0.003</td>
</tr>
<tr>
<td>Faith in others (FI)</td>
<td>0.068</td>
<td>0.011</td>
<td>0.129**</td>
<td>0.118**</td>
<td>0.074</td>
<td>0.080</td>
<td>0.071</td>
<td>0.091</td>
</tr>
<tr>
<td>Self assured (FI)</td>
<td>0.003</td>
<td>-0.048</td>
<td>-0.030</td>
<td>0.019</td>
<td>-0.010</td>
<td>0.000</td>
<td>-0.044</td>
<td>0.039</td>
</tr>
<tr>
<td>Cogn Prefe (NFC)</td>
<td>0.004</td>
<td>-0.026</td>
<td>0.029</td>
<td>0.001</td>
<td>0.001</td>
<td>0.000</td>
<td>0.058</td>
<td>0.057</td>
</tr>
<tr>
<td>Cogn Practice (NFC)</td>
<td>0.010</td>
<td>0.023</td>
<td>-0.032</td>
<td>0.029</td>
<td>-0.033</td>
<td>-0.074</td>
<td>-0.003</td>
<td>-0.116</td>
</tr>
<tr>
<td>Trust in cogn (NFC)</td>
<td>-0.008</td>
<td>-0.026</td>
<td>0.001</td>
<td>-0.015</td>
<td>0.068</td>
<td>-0.022</td>
<td>0.017</td>
<td>-0.040</td>
</tr>
<tr>
<td>D_Gender</td>
<td>0.009</td>
<td>0.043</td>
<td>0.103</td>
<td>-0.064</td>
<td>0.109</td>
<td>-0.025</td>
<td>-0.025</td>
<td>-0.009</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.021</td>
<td>0.015</td>
<td>0.037</td>
<td>0.029</td>
<td>0.026</td>
<td>0.017</td>
<td>0.035</td>
<td>0.032</td>
</tr>
</tbody>
</table>

Notes: * \( p < 0.05 \); ** \( p < 0.01 \)
service responsiveness when they were less involved in the category (\(M_{\text{high}} = -0.130; M_{\text{low}} = 0.194; p = 0.002\)). Further analysis into each item of the empathy and responsiveness service criterion showed two significant areas: listen attentively (\(M_{\text{high}} = 0.106; M_{\text{low}} = 0.137; p = 0.023\)) and demonstration (\(M_{\text{high}} = -0.115; M_{\text{low}} = 0.148, p = 0.014\)), respectively.

Table VIII also shows significant relationships (\(p < 0.05\)) between the MCs’ faith in intuition and the ratings assigned to two service evaluation criteria: empathy and assurance. This indicates that MCs who vary in faith in intuition characteristics rated their service evaluation differently. The MANOVA confirmed that faith in intuition impacted more the ratings on service empathy (0.045) than assurance (0.083). MCs with high faith in intuition assigned higher ratings on service empathy (\(M_{\text{high}} = 0.088; M_{\text{low}} = -0.123\)) and assurance (\(M_{\text{high}} = 0.148, p = 0.002\)).

### 6.2 Theoretical discussions
Theoretically, the capability to perform an MC function is rooted in multiple personas, trait and social roles. Although some of the variability in the way individuals express their personality seems naturally intrinsic, as well as the social role, both non-normative and normative, represents a subsequent need for social integration, the MC job responds to its own logic. Our personality plasticity allows us to flexibly adjust but our social condition forces us to fit by matching others’ expectations in tune with a well-defined role script. This study analyzes the outcome of how individuals manage those two conditions under a specific relevant marketing research activity: evaluating a service. Theoretically, there are two opposing perspectives. Ultimately, the rationalist view is the only one that matters to assure an acceptable return on MC investment.

MC is a useful tool for evaluating real-time marketing service routines. However, concerns on reliability and validity have cast doubt on the legitimacy of the technique for service evaluation. Surprisingly, only a handful of academic content examined this problem (León et al., 1994; Morrison et al., 1997; Wilson, 1998, 2001; Dawes et al., 2000), and even within this contribution, research paucity, loopholes and deficits exist. Our study addresses this gap by investigating the extent to which different MCs would report similar results despite their differing personality. The results suggest that MCs can provide highly reliable data despite the variation in MCs’ individual profiles. However, consideration should be given to MCs’ involvement in product category and faith in intuition. Although the impacts were minimal, the study reinforced the need for a standardized research design to minimize the possible contamination of individual idiosyncratic influences. The findings also suggest that the
agency responsible for the MC must exercise caution in selecting MCs to carry out the evaluation, as well as invest in training strategies that alert the MC of the possibility of personal biases in evaluation. These results also can be useful to scrutinize the researcher’s personal traits on other methods such as participant observer (Bowen, 2008).

6.3 Managerial and social implications
Because of its impact on quality and profitability in high-contact marketing service, these findings have several implications for practitioners. First, it provides empirical evidence that supports the potential subjectivity in the MC evaluation. Our findings show that the ratings on tangible and intangible service factors vary by personal disposition mainly to the extent to which the MC considers him/herself as involved in the category. It shows that the nature of the interaction is perceptual and varies from individual to individual. However, we argue that with instrument structure and employee training, the effect of opinion leadership on service evaluation can be curtailed. For instance, questions and response on the survey evaluation form should be clearly worded and highly structured leaving very little room for ambiguities and interpretation. Second, by alerting the potential MC of the personal tendency to engage in subjective evaluation beforehand, they will be more aware and prepared to recognize and avoid possibility of contaminating the survey. The findings also have implications for the selection of the MC. Recruiters should avoid MCs who are egotistical, or have the desire to individuate themselves from peers. But, recruiters should select MCs with an “open” personality, impartial and neutral. However, the MC should match a specific personality profile (Finn and Kayandé, 1999). Finally, underlying the rational view of marketing services operations, a deviation to the standard behavioral normality is not a “good thing”, as it compromises operational efficiency. However, competitive advantage depends on the ability of service providers to personalize (Surprenant and Solomon, 1987). Similarly, while quality calls for standardization, productivity engagement impacts directly employee’s compensations (Luria et al., 2014). In franchising retailing format, the dilemma arises between standardization and flexibility or adaptation to local specificities (Pardo-del-Val et al., 2014). Those three conflicting goals – efficiency versus personalization, quality versus productivity and standardization versus flexibility – are difficult to reconcile. However it is important to measure exploration effort – standardization, efficiency and quality – and exploitation practices, personalization, productivity and flexibility (Zhanga et al., 2012). A possible solution could be developing two different teams of MCs. One team would match a psychographical profile focusing on the traditional orthodox role of service evaluation. An alternative MC team should be specialized to capture the innovative employee approaches and their strategies to offer more imaginative and personalized services. Using specific psychographic tests, it would be feasible to previously assess MC applicants during the recruiting process and then allocate them to specific tasks.

The scope of the application of MC covers virtually every service activity whether profit- or non-profit-oriented such as hospitality, retailing, health care, banking, etc. The ultimate major goal behind the implementation of the MC method is service improvement and failure correction. Therefore, every step accomplished to contribute to enhance the confidence and accuracy of the data collected is indirectly instrumental to provide quality and social welfare.

6.4 Limitations and future research
Despite the contributions to service marketing literature, this study is not without limitations. First, the study investigated the impact of key personality traits on one service industry – retail computer and electronic sectors. It would be useful to extend this investigation to other service domains like banking, hospitality or transportation by incorporating in a future research the type of services’ variable into the MC service evaluation. Second, the study is limited to examining the differences on homogeneous sample – comprising a restricted age cohort. Additionally, the study only examined the effect of personality traits in isolation of other individual variables that may account for differences in consumer behavior. Our results are contingent to a European cultural background; changing the socio-cultural context may vary the kind, the diversity and intensity of MC personality effects on collected data accuracy (Hopkins et al., 2009).

Future research should consider the effects of demographical factors such as age and gender on the general reliability and validity of MC surveys. As it was mentioned earlier in Section 4.1, an experimental factorial design is an alternative data-collection method particularly useful in fine-tuning detailed analysis of some contextual interference on MCs’ reliability. The MC method is commonly applied in exploitation quality practices, as it evaluates the current well-established “normal” activities. Although more complicated to develop and implement, it would be challenging to measure the exploration/innovative quality practices (Zhanga et al., 2012). Apart from the conventional service quality and job performance assessments, a measurement of employee satisfaction in the perspective of client, as well as the evaluation of the external customer mind-set, would be also interesting to include in the service evaluation instrument (Iyer and Johlke, 2015). We tend to assume that salespeople’s mission is to please customers, but under some situations, they should manage problematic customers and unethical behavior (Suquet, 2010; Madupalli and Poddar, 2014). Such circumstances require specific skills, which also should be analyzed in the MC research process.

References


Further reading


Appendix

Mystery client script

- “Good morning” (greetings).
- “I am interested in a Mp3/4 or digital camera” (Then wait until the employee asks you something).
- “I just want it to download music and videos”, (MP3/4) or “I am interested in a Mp3/4 or digital camera” (Then wait just ask to see more products).
- “I just want to take pictures, no hassles”, (digital camera).
- Always ask to see more products.
- Never express or provide any clue about your preferences or technical specification whatsoever (keep mentioning what is pinpointed in Point 3.).
• After having seen several models/brands – and when, apparently, there is no more left to see – say this: “Which brand or model do you advise me to buy or do you thing fits my goals?”
• Listen his/her answer and then reply: “My cousin (or friend, brother) told me that this brand [. . .] at [. . .]. (competitor retail chain) is cheaper!!”
• Speaking to your colleague (make sure the employee listen the dialogue too) say this “ Do you know what happened to my friend Jo, his Mp3/4 or digital Camera was under warranty time and when it broke down, he brought the device here at the service repair and they complicated everything [. . .]”. Then listen the reaction and do not add more.
• “Is it possible to inform me when a more appropriate model is launched?”
• “Is it possible to let me know when there is a promotion of that brand/model?”
• Ask for a brochure – if they do have it ask a photocopy of technical specifications to analyze it at home.
• Ask the seller for his/her business card.
• “Thanks” (good-by salutation).

Store/salesperson evaluation (attributes)
Note: the digits indicate the corresponding order of the measured attribute shown on Table VI:
1 Assurance (semantic differential seven-point scale):
   • Technically incompetent [. . .] competent
   • Trustful [. . .] untruthful
   • Confuse or muddy in his/her presentation [. . .] clear/enlighten
   • Commitment [. . .] irresponsible/detachment
   • Proactive [. . .] reactive
2 Empathy (seven-point scale: 1/11005 to 7/11005 Applies completely):
   • He/she politely interacted with the customer
   • He/she revealed pleasure/happiness in pursuing his/her job
   • He/she seemed to develop their arguments in an honest/sincere way
   • His/her behavior (verbally and non-verbally) seemed to be ethically correct
   • He/she took seriously and showed interest/concern to help the customer
   • He/she made their best to understand customer’s needs
   • He/she seemed to be selfishly interested in selling instead of helping the customer (r)
   • He/she is truly concerned in helping the client
   • He/she listened attentively what we wanted
   • He/she was serene/calm in finding the solution
3 Tangible evidence – people (semantic differential seven-point scale):
   • Employee well dressed [. . .] dowdy
   • Well shaved/proper make-up [. . .] careless
4 Tangible evidence – place:
   • Dirty [. . .] clean
   • Attractive window-shopping [. . .] uninteresting
   • Easy to find the products [. . .] difficult
   • Messy [. . .] tidy
   • Easy access to products [. . .] physically inaccessible
   • Poor lighting [. . .] good lighting
5 Reliability (semantic differential seven-point scale):
   • Efficient (quick response) [. . .] inefficient (slow)
   • Unconvincing [. . .] very persuasive
   • Offer accurate information [. . .] inaccurate

Corresponding author
Pedro Quelhas Brito can be contacted at: pbrito@fep.up.pt

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
This article has been cited by: