



Mystery Shopping Raised to Scientific Evidence

Mystery shopping is a long-standing method for learning more about what really goes on in the marketplaces of the nation. Now, more than sixty years after mystery shopping first came onto the market research scene, its potential as expert scientific evidence is being recognized. It's all a matter of standards. **Ruth M. Corbin, CMRP**, and **Sarah Carnegie**

From its modest beginnings in the mid-1940s, mystery shopping has reportedly expanded to an established billion dollar industry in North America.¹ Mystery shopping is a method of “market reconnaissance.” It entails collection of information about the customer experience by researchers presenting themselves as customers. Trained researchers visit stores or other retail organizations, send Internet inquiries, or phone for service, posing as customers of the establishments they are contacting. They typically have specific topics on which they are expected to report back to a project sponsor.

Mystery shopping is frequently used by retail and franchise organizations to measure the quality of their customer service and to test for employee integrity at the cash register. The Liquor Control Board of Ontario reportedly sends mystery shoppers into its stores no less frequently than seven times per year per store to assess the shopper experience.²

Regulators have used mystery shopping to test for compliance with customer disclosure requirements, such as the requirements surrounding the issuance of debit cards through financial institutions.³ Canadian lawyers Rebecca Cowdery and Prema Thiele published a description of a mystery shopping study of 40 financial advisors⁴, used to test compliance with the “know your client” rule of the Canadian securities industry. Results of the study showed that many financial advisors failed to collect the appropriate amount of information about the client’s financial situation and risk tolerance.

The Financial Consumer Agency of Canada, an office of the federal government, reports that it periodically undertakes mystery shopping to determine how banks are performing with respect to their consumer obligations under the law.⁵

The Canadian Office of Consumer Affairs (OCA) at Industry Canada has funded various mystery shopping investiga-

tions by independent and not-for-profit organizations for consumer education. For instance, in early 2008, OCA funded an undercover investigation of used car dealerships by the Automobile Protection Association (APA). Posing as consumers looking to buy a used vehicle, APA mystery shoppers visited three dozen used car dealers in Toronto and found various deceptive practices and misrepresentations among a majority of them.⁶ Visits were actually recorded by hidden cameras and later archived on broadcaster CTV's web site.

In short, mystery shopping has been adapted to many marketing, regulatory, and commercial management applications.

Recently, mystery shopping has found its way into litigation evidence, to assess representations to customers made at the point of sale. Although such evidence has routinely been collected by private investigators from specified locations, a mystery shopping study can collect such evidence on a broader scale from a statistically representative sample of retail locations. The use of rigorous statistical sampling, combined with quality controls on consistency and objectivity, allow mystery shopping data to meet the standards of other social science expert evidence.

To be more specific about the standards required to elevate mystery shopping data to a scientific standard, one may consult the recent Canadian Supreme Court decision in *Matte!*. The Supreme Court articulated the three criteria of reliability, validity and relevance. Although the Court made its comments in the context of survey evidence, reliability and validity are universal standards governing all of the social sciences. Relevance to the issues at law is itself an obvious third requirement for any evidence tendered to a court.

It is a straightforward next step in the evolution of mystery shopping to document, for this technique, standards that adhere explicitly to the Supreme Court criteria. Part of the job has already been done in different places. The Mystery Shopping Providers Association is a worldwide organization that sets professional standards for mystery shopping. It originated from a collective interest by its members in "improving service quality." The association publishes a code of ethical practices that includes many guidelines consistent with standards of expert evidence, including guidelines about careful administration, independence of shoppers, and good record-keeping. The International Organization for Standardization (ISO) has its own set of standards – ISO 20252-2006(E) – for quality management of market, opinion and social research, to which some mystery shopping companies subscribe. Those standards are largely captured in the published Canadian industry (MRIA) standards for market research. Quality and consistency of process address the validity of any research endeavour. Any of the sets of standards described above would illustrate an expert's attention to matters of quality control and validity.

A key element of validity, in presenting mystery shopping evidence, is that the mystery shops credibly reflect true customer experience. This element is not explicitly referred to in any of the sets of quality standards discussed in the previous paragraph. It is a matter of plausibility, as assessed by a trier of fact. Validity is enhanced if the mystery shoppers are trained in a thorough, consistent manner and have an opportunity to practise in advance. Validity is also enhanced if the mystery shoppers do not know the underlying purpose of the project and would thus be less vulnerable to bias. Experts advising a court should be prepared to document attention to such validity issues.

As discussed earlier, reliability is another of the tripartite Supreme Court standard for social science evidence. Reliability permits results of a research project based on sampling to be generalized to the overall pertinent population. As with routine surveys, reliability can be reasonably assured by random or systematic sampling, monitoring for consistency, thoroughness of record-keeping, and verification of data coding and data entry.

In summary, there is an opportunity to give triers of fact comfort about the reliability and validity of mystery shopping as expert evidence by collecting published standards from recognized professional associations, or by adapting well-established standards of social science to this particular methodology. Mystery shopping expert evidence need not be regarded as too risky or creatively bold, if courts can be assured that its scientific underpinnings are well-secured.

American case law confirms that "salespersons who influence prospective purchasers" can be reasonably considered to be part of the pertinent population.⁸ But scientific standards of reliability will govern a court's receptivity to mystery shopping evidence as probative or not. In *Tiffany (NJ) Inc. v. eBay*, famous jeweller Tiffany brought an action against eBay for permitting the sale of counterfeit Tiffany silver jewellery on its website. Tiffany's evidence included online mystery shopping, which it referred to as "Buying Programs." Mystery shoppers sought out jewellery for sale on eBay using the search words *Tiffany* and *sterling*. They then proceeded to purchase a random selection of 186 pieces of such jewellery. Upon receipt of the purchases, quality inspection personnel at Tiffany inspected the items and found 73 per cent to be counterfeit. The mystery shopping Buying Programs were rejected by the



court as having little probative value, for reasons that will be familiar to survey experts:

[The] original survey specification suffered from an inherent flaw, namely that it would be impossible for all items to have a known, non-zero chance for selection, and thus that the sample was not a probability sample ... Because the Buying Programs were not probability samples, it is not possible to generalize or make any greater inference about the general population of the sampling study's universe. ... It is impossible to calculate a confidence interval in the data. ... The universe of goods that were sampled through the Buying Programs were not representative of the universe of Tiffany silver jewelry at issue in this litigation.

The court also found several shortcomings in the quality and accuracy of implementation of the study.

While mystery shopping, done to scientific standard, is capable of living up to the basic requirements of reliability that govern routine surveys, there is a special set of exceptions that must be recognized. Mystery shops are simulations of real-life experience. They cannot be, nor should they be, controlled to the same automaton ideal as ordinary surveys. Mystery shoppers must be prepared to interact plausibly with the salespeople, service people, or others whom they encounter. Such an ap-

proach means giving up some measure over control of the situation and being prepared to respond to the unexpected. Quality controls can still be implemented at the level of flexibility that mystery shoppers exercise. For example, if they are unexpectedly asked for more background information than they have, they can be trained to deflect the question.

Social scientific mystery shopping evidence has been tendered in at least two published decisions in Canada. In one case, mystery shopping contributed to the successful appeal in the Federal Court of Canada by the Seiko Corporation (Seiko), manufacturer of SEIKO watches.⁹ Seiko had been turned down by the Trade-marks Opposition Board on its application to register the name SEIKO in association with watch boxes. The Board had taken the position that the boxes were merely packaging material. In order to succeed in its appeal, "it was necessary for the Applicant to prove as a matter of mixed fact and law that a watch box is a ware independent from the watches."¹⁰ Seiko commissioned a research project entailing a broad-based mystery shop of the retail marketplace in Montreal, from which the following conclusion was drawn:¹¹

Retail wares across many product sectors, including watches, may be sold in durable containers. The value perceived by a purchaser in a core product is likely enhanced

by the container it is sold in. The container is not necessarily intended by the seller as, or perceived by the buyer as, part of the packaging. In fact, it may be perceived as a ware in its own right. The characteristics of certain SEIKO watch boxes ... are consistent with containers for other retail products, where the containers are positioned as wares with their own distinct value.

The court found the evidence to be "substantiated by detailed analysis, cogent, consistent and persuasive, and ... deserving substantial weight."¹²

Mystery shopping was used in another case to demonstrate trade-mark confusion among store clerks. Since store clerks can influence hundreds of consumers a day, the research contributed to one party's argument of the likelihood of consumer confusion. The case, heard by the Superior Court of Justice of Ontario, involved a mystery shopping survey tendered into evidence on an interlocutory injunction application by Victoria's Secret against La Senza.¹³ Victoria's Secret had sought to enjoin La Senza and its affiliates from advertising and selling its ITEC brand bra. Victoria's Secret felt that the name ITEC, combined with elements of its advertising campaign, would cause consumer confusion with Victoria's Secret's IPEX brand bras.

The mystery shop consisted of contacting every one of the defendant's stores



in Canada, asking for the plaintiff's brand. More than two-thirds of the clerks contacted said they did have that brand, or expected getting it within the next few weeks. The court found that the mystery shopping study provided ample evidence of a serious issue to be tried.¹⁴

Mystery shopping holds further potential as an innovative purveyor of relevant expert evidence. Its advantages shine when there is insufficient time for traditional consumer surveys, or when the complexities of a consumer survey present too many barriers. But mystery shopping studies need not be second choice to traditional surveys. They have several superior qualities, including the direct simulation of real life. Their ability to "approximate real market conditions"¹⁵ was noted with approval by the court in *Lon Tai Shing Co. v. Koch + Lowy*, in which mystery shoppers visited lamp stores and showed salespeople a photo of a lamp they were interested in purchasing. The picture showed either the plaintiff's lamp, the defendant's lamp, or a control lamp. As for specific evidentiary value, mystery shopping studies have proven suitable for measuring, for example, the likelihood of confusion among consumers (when influence by clerks is at play as in *Victoria's Secret v. La Senza*, above), evidence of standard commercial practice in particular retail sectors (as in *Seiko*, above), and the presence of mis-

representation, intentional or otherwise (as in *Tiffany v. eBay*, above, had sufficient standards of reliability been met). Other possible applications are limited only by the imagination of counsel, or by available budget. Mystery shopping is a recognized established method of data collection, capable of incorporating standards of validity and reliability, and ready for expanded applications in expert evidence.

REFERENCES

1. Reported on CBC Broadcast *Marketplace*, January 9, 2001. Researcher James Donne, reporter Jacque Perrin, producer Richard Wright.
2. Reported by LCBO directly, at www.lcbo.com/aboutlcbo/todayslcbo.shtml
3. In 1995, Industry Canada and the Canadian Bankers Association commissioned CorbinPartners Inc. (then Decision Resources Inc.) to conduct a mystery shop of debit-card-issuing financial institutions, to assess compliance with the 1992 *Canadian Code of Practice for Consumer Debit Card Services*. Following on the release of the results, the code was revised in 1996.
4. R. Cowdery & P. Thiele. "Knowing Your Client Is a Business Necessity," in *Advisor's Edge Report*, 2007 (Oct.), p. 7.
5. Financial Consumer Agency of Canada, Background - 2004-05 FCAC Mystery Shopping Results (undated). Accessible at www.fcac.gc.ca/eng/Publications/SurveyStudy/Myst-Shop/PDFs/MysteryShopping2004-2005_e.pdf (last visited on November 26, 2008).
6. A more detailed summary of results can be found at www.apa.ca/template.asp?DocID=136 (last visited November 26, 2008).
7. *Mattel, Inc. v. 3894207 Canada Inc.*, [2006] 1 S.C.R. 772, 2006 SCC 22, at paragraph 45: "As to the usefulness of the results, assuming they are elicited by a relevant question, courts have more recently been receptive to such evidence, provided the survey is both reliable (in the sense that if the survey were repeated it would likely produce the same results) and valid (in the sense that the right questions have been put to the right pool of respondents in the right way, in the right circumstances to provide the information sought). See *Canadian Schenley Distilleries Ltd. v. Canada's Manitoba Distillery Ltd.* (1975), 25 C.P.R. (2d) 1

(F.C.T.D.), at p. 9; *Joseph E. Seagram & Sons Ltd. v. Seagram Real Estate Ltd.* (1990), 33 C.P.R. (3d) 454 (F.C.T.D.); *Walt Disney Productions v. Fantasyland Hotel Inc.* (1994), 20 Alta. L.R. (3d) 146 (Q.B.). Thus, in *Cartier Inc. v. Cartier Optical Ltd.* (1988), 20 C.P.R. (3d) 68 (F.C.T.D.), the court accepted as helpful a survey found to be properly designed and impartially administered and whose findings were directly relevant to the likelihood of confusion. This was also the case in *Sun Life Assurance Co. of Canada v. Sunlife Juice Ltd.* (1988), 22 C.P.R. (3d) 244 (Ont. H.C.J.)."

8. *I.P. Lund Trading APS v. Kohler Co.*, 118 F. Supp. 2d. 92 (D. Mass. 2000) at 106.
9. *Kabushiki Kaisha Hattori Seiko v. Impenco Ltd.* (1999) Federal Court, (November 03, 1999), Docket number: T-1559-98.
10. See reference 9: paragraph 4.
11. See reference 9: paragraph 8.
12. See reference 9: paragraph 9.
13. *Victoria's Secret Stores Brand Management, Inc., et al. v. La Senza Inc.*, Superior Court of Justice - Ontario, Court File No. 06-CV-306933 PD1, March 30, 2006.
14. See reference 13: paragraph 9.
15. *Lon Tai Shing Col. v. Koch + Lowy*, No. 90 Civ. 4465 1991 WL 170734 (S.D.N.Y. June 20 1991).

Corbin Carnegie on Mystery Shopping, copyright protected

Dr. Ruth M. Corbin is chief executive of CorbinPartners Inc., adjunct professor at Osgoode Hall Law School, and co-author of the recent Butterworth text Survey Evidence and the Law Worldwide. Ms. Sarah Carnegie is project manager at CorbinPartners Inc., and PhD candidate at the University of Calgary. They may be reached at (416) 413-7600 or by email at rcorbin@corbinpartners.com and scarnegie@corbinpartners.com