Marketing Research in the Digital Age

Raymond R. Burke
Arvind Rangaswamy
Sunil Gupta

April 1999
Revised December 1999

Raymond R. Burke is the E. W. Kelley Professor of Business Administration and director of the Customer Interface Lab, Kelley School of Business, Indiana University, Bloomington, IN 47405. Tel: (812) 855-1066. Fax: (812) 855-6440. E-mail: rayburke@indiana.edu.

Arvind Rangaswamy is the Jonas H. Anchel Professor of Business Administration and the Research Director of the eBusiness Research Center, The Smeal College of Business Administration, The Pennsylvania State University, University Park, PA 16802. Tel: (814) 865-1907. Fax: (814) 865-3015. E-mail: arvindr@psu.edu.

Sunil Gupta is Executive Vice President, Acorn Information Services, 4 Corporate Drive, Shelton, CT 06484. Tel: (203) 225-7600. Fax: (203) 225-7610. E-mail: sgupta@acornis.com.

Abstract

Marketing Research in the Digital Age

In the past few years, digital technologies have stimulated several innovations in marketing research. These innovations include online survey methods and focus groups, e-commerce customer tracking systems, data mining tools, and 3-D graphics software. Technology is changing the way marketing data are collected, analyzed, and used for supporting managerial decisions. In this paper, we provide an overview of these innovations, discuss their benefits and limitations, and consider how these developments will alter the nature and scope of the marketing research function in the future.

1. Introduction

Marketing research provides managers with information that can be used to identify and define marketing opportunities and problems; generate, refine, and evaluate marketing actions; monitor marketing performance; and improve our understanding of marketing as a process (adapted from the definition provided by the American Marketing Association, Marketing News, January 2, 1987).

The theory and practice of marketing research are on the verge of a major transformation brought about by the growing deployment and application of digital technologies, including point-of-sale UPC scanners, frequent shopper programs, credit/debit cards, in-store tracking, caller-id systems, virtual reality simulations, and most importantly, the Internet (Kannan, Chang and Whinston 1998; cf. Malhotra, Peterson and Kleiser 1999). These technologies have produced an explosion of data on customers and their purchase behavior. Online data collection, storage, and retrieval systems deliver volumes of secondary data on market trends and the competitive environment. Syndicated databases now provide detailed information on the buying habits of individual households. Online surveys have simplified and accelerated primary data collection. Virtual reality simulations now allow marketers to test new marketing ideas quickly, inexpensively and confidentially. The task of analyzing marketing data is no longer relegated to research specialists. Increasingly, managers are using networked desktop computers, data analysis and data mining software, and Internet search engines to access and process marketing research information.

We are just starting to see the impact of these changes on the marketing research discipline. As the Internet penetrates more homes and businesses, as the bandwidth of the communication pipeline increases, and as the network attaches to new types of devices (including hand-held and wireless devices, kiosks, automobiles, appliances, and RF and video surveillance systems, among others), researchers may be able to collect detailed and current information about customers, their needs and preferences, and their shopping habits. Few aspects of marketing research will be left untouched by the digital medium.

The objectives of this paper are to provide a framework for understanding current developments in digital technology, discuss the benefits and limitations of these approaches, and explore some long-term consequences for the practice of marketing research.

2. Technology's Role in Marketing Research

Traditional marketing research consists of several stages: (1) framing the research problem, (2) developing an overall research strategy, (3) designing data collection methods, (4) collecting, analyzing and interpreting research data, (5) recommending courses of action to resolve the decision problems, and (6) participating in the implementation and evaluation of selected actions. Digital technologies impact all stages of marketing research. However, we will primarily explore the third and fourth stages, which have been most affected by the digital medium.

In traditional marketing research, it is common to categorize research data collection into two types: primary and secondary. Primary data are collected for the purpose of a specific research project. This involves obtaining data directly from respondents. Secondary data are data that already exist, collected for reasons other than the specific project of interest. Digital technologies, particularly the Internet, have significantly increased the quantity and accessibility of secondary data and expanded our capability to collect primary data. We explore these two developments in greater detail below.

Secondary Research

In the physical world, secondary research has the reputation of being limited in scope, lacking in detail, and often out-of-date. In the digital world, the reverse is true. The Internet provides volumes of useful and timely information, much of it for free. The fact that the Internet is a global phenomenon makes it an even more valuable source of secondary data. In many cases, information about distant markets is easier to get online than by any other means.

A key benefit of conducting secondary research on the Internet is that much of the online information has been indexed and is searchable. In some cases, this index is constructed by a program called a "spider" which searches the web and electronically catalogs the information on each web page. In others, web directories are compiled manually. Users can search these indexes using a general-purpose search engine, such as those found at altavista.com, excite.com, hotbot.com, yahoo.com, and looksmart.com. There are also search engines with unique features, including MetaCrawler (metacrawler.com), which can simultaneously search using multiple search engines, Google (google.com), which is designed to retrieve the most relevant Web information, and Northern Light (northernlight.com), which searches over 120 Million Web pages and a special collection of more than 5,400 full-text sources. (For additional information about search engines and their relative merits, readers should visit the web site searchenginewatch.com.)

The cost of search technology has dropped rapidly in recent years. It is now feasible for a web site to provide its own search engine to help users find information available only at that site (e.g., New York Times – nytimes.com). There are also search engines ("agents") that can collect information from a specified set of participating sites and dynamically create "virtual databases." Price search agents (e.g., Junglee, now owned by amazon.com) represent one such category of search engines.

HTML, XML and related standards provide uniformity to electronic documents, making it possible to access and display a variety of information through a single, familiar interface – the browser. Even such traditional purveyors of information as Lexis-Nexis and the U.S. Census Bureau have made their databases accessible through the Internet. Lexis-Nexis contains the full text of 22,000 major publications—magazines, newswire reports, and even TV program transcripts—dating back several years. In addition to standard search techniques (e.g., using Boolean operators, such as "and" "or" and "not" in search strings), Lexis-Nexis allows the researcher to do positional searches of full-text documents (e.g., online "within 10 words" of marketing research; Venetian "before" blind, rather than blind "before" Venetian). Soon, most

syndicated data sources (including those provided by firms such as IRI and A.C. Nielsen) will be accessible through the Internet.

Here are a few examples of online sources of secondary data useful for marketing managers:

- Company information: Detailed company information is available from several Internet web sites. A good source of information about public companies is the U.S. Securities and Exchange Commission (sec.gov). The SEC has set up the EDGAR database (the Electronic Data Gathering, Analysis, and Retrieval system), which archives most forms filed by public companies. Its primary purpose is to increase the efficiency and fairness of the securities market, but it is also a valuable source of marketing data. For example, by searching the database for a company's most recent form 10-K, you can often find more detailed information than what is reported in the company's annual report. Another government source is the U.S. Patent and Trademark Office (uspto.com). The USPTO database allows you to search for patents and copyrights held by a company, which can suggest a firm's future business directions.

  There are also several private sources of company information, including the Wall Street Research Net (wsrn.com), Company Sleuth (companysleuth.com) and Hoovers Online (hoovers.com). In most cases, these services provide free access to recent news stories and other public information about companies, but charge for more detailed reports based on proprietary research studies. Another helpful site is CEO Express, which offers convenient access to most business magazines and news sources (ceoexpress.com).

- Consumer information: There are also a growing number of sites which provide detailed consumer information. The mother lode of consumer data is the U.S. Census. Through the U.S. Census Office's web site (census.gov), you can access information on over 60 individual and household characteristics, including age, education, income, occupation, education, age and value of home, telephone availability, etc. The data are available for the entire U.S. population, aggregated from the individual block group, up the census tract, zipcode, county and state levels. The U.S. Bureau of Labor and Statistics (stats.bls.gov) provides up-to-date information on employment. Many of the traditional syndicated services, such as Mediamark (mediamark.com) Arbitron (arbitron.com), and USA Data (usadata.com) allow you to download consumer data directly from their web sites.

- Information about the Internet: As an important and growing medium, the Internet itself is a popular subject of research. Information about the Internet and its various uses is available from several web sites, including cyberatlas.com and iconocast.com. These sites provide free access to summaries of recent research reports on the demographic and psychographic profiles of Internet users, their online shopping habits, the amount of money spent on web advertising, and information on a host of other aspects. Likewise, Internet companies such as Netscape and Amazon.com, provide a substantial amount of up-to-date information on their web sites describing their operations.

Consumer evaluations of products and web sites are also available at several sites, including e-buyersguide.com, e-pinions.com, and amazon.com. While these sources may provide some indication of consumer sentiment towards certain products and services, researchers should be wary, as the information is collected from a convenience sample of online shoppers and is self reported. In some cases, consumer evaluations are mixed in with paid editorial messages promoting the item.

Another source of online secondary data is the archives of news groups. The Internet is home to numerous interest groups in a wide range of areas, such as new product development, gardening, movies, and amateur astronomy. Discussions between members of these groups are typically archived and available for later access by members. Some companies run their own forums, and do not share all of the information they record with their members. For example, Adobe (adobe.com) is able to capture detailed user information from its support site, including, for example, the number of people who visited the Photoshop forum, the profiles of the most frequent visitors, and their comments about Photoshop and competing products. While some newsgroups have constraints on who can become a member (e.g., Interval Research's Purple Moon site, purple-moon.com, which is targeted at pre-teen girls), most sites have few restrictions (e.g., webMD.com). Marketing researchers who become members of these groups can often get valuable information about market trends, consumer opinions, and emerging product problems.

There are several limitations associated with conducting secondary research on the Internet. General purpose search engines and web directories vary greatly in terms of their coverage of the web and the "freshness" of their links (Lawrence and Giles 1999). Even the best search engines catalog only a fraction of the available information. Online research can also be time consuming and can generate volumes of irrelevant information, especially if the user chooses the wrong search engine for a particular purpose, or does not use efficient search techniques.

To improve the efficiency of the search process, some engines attach an "index" value (e.g., on a scale of 0 to 100) to each of the sites retrieved to indicate how well the information at

a site matches the information requested by the user. Web directories like Yahoo! do an extensive analysis of every web site before including the site's URL in their database. This increases the hit rate of relevant sites included in the search results, but it can be detrimental when one is doing "exploratory" research. Some search engines (e.g., hotbot.com) list search results in the order of frequency with which their users have visited the sites. This is an interesting new application of marketing research – the firm uses information on search effectiveness, generated at its own site, to improve the product it offers. This strategy is akin to Amazon.com's use of "collaborative filtering" technology to recommend books to customers based on the shopping habits of other customers with similar tastes.

Another danger of online secondary research is that the ease of information retrieval may distract users from focusing on the quality of the information. Anyone can publish on the Internet, and users must be skeptical about the validity of online information unless it comes from a known and trusted source.

Primary Research: Self Report Measures

A major reason for the growing interest in online primary data collection is that it is often faster and cheaper than offline research. The primary data collection methods typically used in traditional marketing research are: (1) in-person interviews, (2) focus groups, (3) telephone surveys, (4) mail surveys, (5) observation methods, and (6) lab and field experiments. The advantages and limitations of these methods are well documented (see, for example, Malhotra 1993). All of these approaches have been adapted for use on the Internet. (The only exception is online personal interviewing, which will become practical if and when videoconferencing achieves significant penetration in the home market.) Online surveys using e-mail and web forms, focus groups, experiments, and observation methods have now been in use for several years. Observation methods have become especially popular, as many web sites track the online activities of their customers.

In early 1999, Acorn Information Services and the Institute of International Research (IIR) initiated a longitudinal survey of IIR members to assess their use of online marketing research. Two surveys were administered six months apart. In the first wave, 196 out of the 2,000 members polled responded; in the second, 412 out of the 4,000 members polled responded.

The results indicated that the majority of IIR members had tried online research, using it for the same kinds of applications as conventional research methods (see iir.acornis.net for the full report). Most users conducted relatively simple studies, such as e-mail and web-based surveys for measuring brand attitudes, customer satisfaction, and segmentation. They were less likely to use online research for conjoint analysis, tracking studies, and online discussion groups. The study also revealed that:

- Although there was widespread use of online research by survey respondents, the online budgets were limited. Almost 80% of respondents in the second wave of the survey reported that online research accounted for less than 10% of the total marketing research budget during the past 6 months. However, 39% of these respondents expected the online share of the budget to exceed 10% in the coming year.

- Twenty nine percent of the respondents in the first wave of the survey indicated that online marketing research was either very useful or moderately useful for *most* of their research problems. An additional 47% indicated that online marketing research was either very useful or moderately useful for *some* of their research problems. These percentages did not change significantly in the second wave of the survey.

- IIR members reported that the most important reasons for considering online marketing research were its ability to evaluate technically complex products and stimuli that would be difficult to explain with just text, and to reach hard-to-get respondents.

- They felt that the biggest problem with online marketing research was the lack of sample representativeness. The most important benefits were the lower costs per respondent and the increased speed of data collection.

In the remainder of this section, we discuss the various types of online primary data collection methods.

Online surveys. Perhaps the most common type of Internet-based data collection is the online survey. In its simplest form, the survey is conducted by posting a form containing a set of questions to a site's web server. Consumers who happen to visit the site or are directed to the

site's URL through a banner ad, e-mail message, or other communication, are asked to respond to the list of questions. The responses are automatically captured and stored in a database for later analysis.

Many firms have begun to experiment with online surveys, and a few have compared the performance of online and traditional survey methods:

- Socratic Technologies conducted a side-by-side comparison of telephone and Web-based research for a software company. Overall, the findings from the two surveys were very similar. The results for awareness, use of features and functions, importance of features and functions, interest in new features, and general demographics were almost identical. However, there were some differences in repurchase intent and satisfaction. Web-based respondents tended to have more extreme opinions regarding satisfaction. (Socratic reports that this is a general tendency among online respondents.) For purchase intent, Web-based respondents were less likely to give yea-saying answers, and their responses were more closely related to actual behavior. Socratic noted that online methods allowed the client to collect more data than offline techniques with the same budget. In fact, their marginal cost for online data collection was almost zero!

- Quaker Oats conducted a study with M/A/R/C Online for a rice snack. The company had been doing traditional awareness, attitude and usage studies for the past 9 years, but was evaluating whether to move these studies online. Therefore, they conducted an online survey along with a mall intercept study to maintain comparability with earlier results. Online data were collected using AOL's Opinion Place. An elaborate screening system was designed to qualify respondents, who were awarded 250 points (with a monetary value of about $2.50) for their participation. The demographics of the online sample mirrored that of the mall sample. The results were essentially the same in the two samples. There were, however, more "Don't Know" responses online. ("Greater frankness" is how some marketing researchers have explained this recurring result.) The company was able to survey three times as many people online as with the mall sample for the same budget. The additional sample size allowed the firm to examine sub-segment level behavior in greater depth.

- A high-tech toy company conducted a web-based concept test and conjoint study. The company compared the web survey results with the more conventional disk-by-mail (DBM) method. The response rate was higher for DBM. There were nearly identical results in the two samples regarding features assessments and the perceived skills that result from using the toy. Online respondents had lower purchase intent scores (less yea saying?) than DBM respondents. However, across concepts, the rank orders were highly correlated. The conjoint part-worths were also highly similar, as were simulated market shares for the concepts. One difference was a slightly higher price sensitivity for the Web-based sample. The web surveys were faster, afforded greater flexibility, and were much cheaper.

As we can see from these examples, the cost per respondent of online research is typically lower than offline research. By distributing the questionnaire electronically, we reduce production costs and eliminate mailing and/or interviewing costs. Because of the convenience of taking an online survey, lower respondent incentives are necessary. Web surveys can be conducted in 20 days or less at a cost of between $10,000 and $20,000. The exact price depends, on the size of the sample and the complexities of skip patterns and question rotations. In Exhibit 1, we compare the costs and benefits of conducting a concept test using online and offline marketing research techniques.

Online research is also faster. Consider the case of a print magazine for teenagers that wanted to test alternative cover concepts quickly. The publisher selected 300 respondents, matching their readership criteria, from a research company's online panel. Respondents were sent an e-mail and invited to visit a Web site where they saw and evaluated the various concepts. Responses started arriving within 15 minutes of mailing the participation request. Most of the responses were received within 24 hours. (Several other researchers have reported similarly rapid response rates.) The client then made changes to the cover, and once again invited the respondents to score the concepts. This sequence of feedback was repeated 3 times in 5 days, with a total research cost of under $6,000.

The online medium is particularly suited for certain types of survey research that can take advantage of the interactivity and tracking ability of the medium. For example, bizrate.com has developed technology which allows companies to conduct point-of-sale customer satisfaction surveys with a randomly selected group of respondents immediately after they complete a (purchase) transaction at an affiliated site (Reibstein 1999). Compared to traditional research methods, online research also allows companies to reach a more geographically dispersed sample of respondents.

A major disadvantage of all online surveys is that the sampling frame is limited. Further, there is no master list of all people with Internet access, as there is with telephone numbers. (Note also that there is no online equivalent to Random Digit Dialing.) Most people still do not

have Internet access (current estimates are around 25% of the population). Thus, a major concern with online research is that Internet users may not be representative of the general population. For example, Internet shoppers have traditionally been males with higher than average levels of income and education (see, for example, Reibstein 1999). While this group may be appropriate for research on some topics, such as web site content, purchases of technologies and computer products, it does not match the target demographic of most consumer products and services.

Fortunately, as Internet penetration increases, the characteristics of Internet users are becoming more representative of the general population. Low priced PCs are being purchased by a growing percentage of lower- and middle-class households (Briones 1998). IntelliQuest reported that the percentage of online users without a college education jumped from 54 percent in the second quarter of 1996 to 64 percent in the first quarter of 1999. Online households with average incomes below $50,000 increased from 40 percent to 45 percent during the same period. The Pew Research Center reported that 23 percent of those who have gone online in the previous year make less than $30,000 per year. According to a March 23, 1999, report by the Yankee Group, the household Internet penetration rate in the U.S. is now at 25 percent, and should be close to 67 percent by the end of 2003. Overall, it appears that differences between the online and offline populations are narrowing (see, for example, the sequence of GVU survey results at cc.gatech.edu/gvu/user_surveys).

Because the chances of getting non-representative samples are high online, researchers have developed several ways to address this problem. This includes the following tactics:

- *Select samples from a panel*: Many online research firms, such as NFO Interactive, NPD, Greenfield Online and Cyber Dialogue, have established large panels that have been pre-selected to be representative of the online population. These research companies can select samples from these panels to meet the needs of specific studies. Another advantage of panels is that the typical problem of online surveys, namely, "respondent authenticity," is no longer an issue. However, panels cost money to maintain, and this reduces some of the cost advantages of online research.
- *Adjust online research results:* Some research firms are developing proprietary weighting schemes to adjust online research results to account for known demographic differences

between the online sample and the composition of the desired target population. This might, for example, involve over-weighting the responses of females and under-weighting the responses from males. The appropriateness of these weighting schemes has not yet been evaluated carefully.

- *Pre-qualify respondents through e-mail solicitation*: This reduces self-selection bias, which occurs when web surfers happen to come across a web survey and choose to fill it out[1]. Further, e-mail solicitation costs only a fraction of mail solicitation. However, unsolicited e-mail is not viewed favorably by the online community. As a result, this may not be a viable long-term option to improve the representativeness of samples.
- *Put links to the web surveys at popular Internet portals (e.g., Yahoo, America Online, etc.):* While this is likely to ensure a good sampling frame, there is still the problem of response bias. The people who choose to click on these links are likely to be systematically different form those who do not.[2] Further, not everyone who comes to a survey site through a clickthrough will actually complete the survey.
- *Offer incentives for participation*: In the online medium, it is particularly easy to administer incentive programs, such as frequent flier miles. For example, respondents can be offered points or free gifts for participating in the study. However, without proper safeguards, it is also easier for respondents online to misuse an incentive program by repeatedly filling out the surveys.

In spite of the above tactics, there will always be questions about the representativeness of online samples until most people have online access. It is wise to remember that it took more than 30 years after the first telephone interviews were conducted for it to become a common research method. Nevertheless, for certain topics (e.g., evaluation of a web site), online surveys are likely to generate representative samples and have no more disadvantages than the "mall intercept" study done in traditional marketing research.

Online focus groups. A focus group, or group interview, is an exploratory research method used to help companies gain a deeper understanding of customers' needs and wants, shopping habits, decision criteria, and reactions to new marketing ideas. Focus groups bring together six to twelve participants in an informal setting to discuss their attitudes, perceptions, motivations and desires. An interviewer moderates the discussion based on an outline provided

[1] A web survey is designed using HTML or XML and is posted at a web site. For examples of such surveys, visit: gvu.gatech.edu/user_surveys.
[2] Clickthrough rate for banners and other links is less than 1% even at major portal sites (see, for example, surveys by NetRatings).

by the client. The group discussions usually last from one to two hours. The discussion is audio- or videotaped, and then transcribed, coded and analyzed at a later time.

Focus groups have several benefits. They can be used for collecting in-depth, qualitative research in a short period of time and at a relatively low cost. It's possible to use physical props and audio-visual materials when testing consumer reactions to new products, packaging, promotions and advertisements. Focus groups allow clients to interact with current and prospective customers, tailoring the line of questioning based on the flow of the discussion. The videotaped interaction provides information on both the verbal and non-verbal reactions of participants.

Focus groups also have limitations. They are typically based on a convenience sample of respondents, so the results cannot be generalized to the population. Another concern is that the dynamics of the group discussion can bias the responses of individual participants.

Two digital technologies are having a significant impact on online focus group research. The first is videoconferencing. Rather than having the client travel to a focus group facility and observe the discussion firsthand, the session is broadcast to a remote facility (often the client's office) where people can observe the live proceedings over a television. This saves clients the time and cost of traveling to distant locations, and allows them to participate in a greater number, and more geographically diverse set, of focus groups. Unfortunately, the picture quality of videoconferencing is inferior to face-to-face interaction because of limited resolution and time delays in signal transmission. Further, the hardware and software costs for group videoconferencing can run into the tens of thousands of dollars. Long-distance telephone charges for ISDN connections are several times more expensive than audio connections. Finally, videoconferencing limits the degree of interaction between the client and the focus group moderator.

The second technology transforming focus group research is Internet "chat" software. Instead of coming to a physical focus group facility, participants join an online group by visiting a secure web site using their home or office computer and a web browser. The chat software

allows the focus group moderator and each participant to type messages into a scrolling window and see the responses of the other participants. People are typically recruited from selected Internet locations, such as newsgroups and forums, but can also be enlisted through online banner ads or telephone solicitation.

Online focus groups have several benefits. They are convenient for both clients and respondents, saving travel time and cost. The chat software automatically logs the interaction, permitting immediate distribution and analysis of the data. Online focus groups are very effective in bringing together people from different parts of the country, especially those in higher income brackets, and are particularly well suited when the subject is personal or sensitive (Weissman 1998). However, online focus groups also have limitations. An important limitation is that the technology limits the interaction between respondents. The chat software typically provides a text-only interface, so it is not possible to show props, listen to voice inflection, or observe the facial expressions and body language of respondents. Typing skills are necessary (although respondents are asked to ignore typographical, grammatical and spelling errors).

In the future, as video cameras, full-duplex (simultaneous two-way) audio, and high-bandwidth Internet connections become common in home computers, we are likely to see more sophisticated versions of online focus groups, providing the richer interaction of videoconferencing with the convenience and low cost of chat software. It is also likely that this technology will incorporate multimedia and 3-D capabilities, so we can measure respondents reactions to new advertising, promotion, product and packaging concepts, as in conventional focus groups.

Primary Research: Experimentation

The Internet is not the only aspect of the digital environment that is changing marketing research. Other innovations, such as 3D graphics, kiosks, and infrared tracking systems are also

increasingly being deployed in improving marketing research methods. In this section, we assess the value of these technologies for marketing experimentation.

Computer-simulated test marketing. One of the areas where technology has improved both the efficiency and effectiveness of marketing research is in test marketing. Test markets are often used at the final stages of product development to provide accurate estimates of consumer demand. They can also be used to evaluate new package designs, pricing strategies, advertisements, promotions and merchandising. While conventional test markets are slow and expensive to conduct, innovations in 3-D computer graphics and virtual reality have made computer-simulated test markets practical for a wide range of applications.

Consider a new product test as an example. In a conventional test market, a company would produce a sample run of the new product, introduce it into a typical midsize market, and track sales using point-of-sale scanner data, warehouse withdrawals, and/or product shipments. This methodology has several benefits. It tests the new product in a realistic competitive context with a representative group of consumers and provides objective measures of marketing performance (e.g., sales and market share). Unfortunately, such research often takes six months to a year or more to complete and can cost millions of dollars. Also test markets can evaluate only a limited number of options and can be monitored and disrupted by competitors.

This has led some firms to search for an alternative approach that provides the realism and accuracy of a test market with the speed, flexibility and low cost of laboratory research. A variety of Simulated Test market (STM) services have been developed, including Assessor, BASES, Discovery and ESP. While these services differ in terms of sampling, questionnaires and modeling, most share a common test procedure of exposing consumers to a new product or marketing program in a simulated shopping environment. In some methods, the manufacturer's product and competitive offerings are shown in a photograph or slide presentation. In others, consumers shop from an actual shelf fixture stocked with products. In either case, the creation of the simulated store is cumbersome, time consuming and costly, especially when the simulation involves a large number of products, categories, manipulations and/or data-collection sites.

When the stimulus presentation does not match the physical store, category-specific calibrations must be applied to the forecasting models to achieve acceptable levels of prediction.

In the virtual shopping simulation, 3-D computer graphics are used to recreate the appearance of a grocery shelf on touch-screen monitors or head-mounted displays (Burke 1996; Burke, Harlam, Kahn and Lodish 1992). Shoppers can pan down the aisles of the store using a 3D trackball, “pick up” products by touching their images on the screen, and rotate packages and magnify labels for closer inspection. To purchase a product, the consumer touches an image of a shopping cart and the package then “flies” into the basket.

Simulations offer several advantages over existing methodologies. They provide the realism and visual clutter of an in-store experiment while delivering the control and process tracing measures of laboratory research. The computer unobtrusively records the amount of time consumers spend shopping in each category, the items they pick up, the amount of time taken to examine individual packages and labels, as well as the quantity of items purchased. Consumer behavior in the virtual shopping simulations has been found to closely mirror behavior in the physical store (see Burke 1996; Burke et al. 1992). Since the new products are being simulated by the computer, testing can occur at an early stage of the new product development process, before manufacturing costs have been incurred.

Unlike most other laboratory research techniques, including focus groups and concept tests, the virtual shopping simulation does not force customers to examine the new product, package, promotion or merchandising being tested. Instead, shoppers can selectively attend to the products and promotions they are most interested in, just like in a conventional retail store. The importance of selective attention becomes apparent when one compares data from a virtual shopping simulation with data from a traditional concept test (respondents rate their intent to purchase on a 5-point "not at all likely to purchase" to "definitely would purchase" scale). Using data from 22 new product tests, the first author found that the concept test scores (top two box) had a significant, positive correlation with the consumer's likelihood of buying a product once he or she had picked it up from the simulated store shelf ($r = 0.58$, $p < 0.005$). However, the

concept test scores indicated no significant relationship with the consumer's likelihood of picking up the product in the first place ($p > 0.10$). It appears that while concept test scores can give a good indication of the appeal of a concept once consumers have noticed the product, they do not predict the product's ability to stand out from the visual clutter and attract the consumer's attention.

Virtual reality simulations have certain limitations when compared to conventional test markets. The simulation can do a reasonable job of replicating the visual and audio information that might be available at the point of purchase. However, if consumers are using other cues to make selections, including tactile (e.g., weight, texture, firmness), olfactory, or taste information, then the simulation will not accurately predict consumer behavior in the physical store. It is especially important to note that there is no physical product consumption in the virtual shopping simulation. Therefore, the simulation can provide an estimate of product trial, but not repeat purchase. Some companies have provided consumers with product samples between shopping trips in order to measure repeat purchase rates.

It should also be noted that, at the present time, virtual reality simulations are novel to consumers. Therefore, consumers initially spend more time navigating through the aisles and interacting with products in the simulation than they would in the conventional store. To overcome this limitation, researchers often ask consumers to go on multiple shopping trips through several product categories. By the third or fourth trip, most consumers are taking about the same time to shop as in the conventional store, and that is when shelf displays are manipulated.

We know of at least 130 commercial studies in which virtual shopping simulations have been used. One of the first applications of the shopping simulation was conducted by the Goodyear Tire and Rubber Company (Rickard 1993). Goodyear was planning to extend its tire distribution beyond the traditional Goodyear stores to include general merchandise and mass merchandise stores. The company wanted to evaluate the equity of the Goodyear name in these new channels and identify the most profitable pricing and warranty options. The research asked

over 1,000 respondents to shop in a series of simulated tire stores where the prices and warranty levels of the various brands were systematically manipulated. The simulation allowed Goodyear to estimate the price and warranty self- and cross-elasticities for all major brands of tires.

Many of the simulation studies have been conducted on consumer packaged goods. For example, Frito-Lay, has used the virtual shopping simulation to examine the impact of end-of-aisle displays and on-shelf advertising on the sales of snack foods. It has also simulated a vending machine to measure consumers' likelihood of switching between salty snacks and candy bars, cookies, crackers and fruit under reduced assortment conditions. Johnson & Johnson's advanced care products division has used the shopping simulation to help price Uristat, a pain killer used for urinary diseases (see Feder 1997).

New technologies will allow marketers to create even more realistic virtual shopping environments. For example, the CAVE is a room-sized virtual reality environment using four video projectors and stereo glasses to create a completely immersive, stereoscopic simulation for one or more consumers. The cost of this technology is currently in the hundreds of thousands of dollars, but is dropping quickly.

Kiosks and electronic shelf labels. While advanced 3-D graphics allow us to conduct computer-simulated test markets in the laboratory, innovations in retail technology are making it easier to conduct experiments in the physical store. For example, the Swedish firm Pricer has developed wireless, electronic shelf labeling which uses liquid-crystal displays (LCDs) to show the names and prices of merchandise. The information displayed on the signs is electronically transmitted from the store's central computer, which also feeds prices to the checkout registers. Therefore, the prices shown on the shelf tags are always consistent with the prices scanned at the register. Since the prices are controlled electronically, they can be changed quickly, allowing retailers to conduct pricing experiments.

Kiosk technology is also being used to test the performance of marketing programs. For example, Interact Systems has installed touch-screen kiosks in several major retail chains. When a shopper inserts a "frequent shopper" card into the kiosk, it displays a customized set of

products and store specials based on the customer's past purchases. Different offers can be targeted at matched groups of customers. When the shopper presents his or her frequent shopper card at the checkout counter, the system automatically deducts the promotional discount from the order. The kiosk and checkout registers are linked, so we know what percentage of the shoppers who were exposed to an offer went on to redeem it.

Primary Research: Observation

A web presence allows marketers at a firm to track what visitors do at their site. At the minimum, all web servers allow measurements that conform to the "common log format." Every activity of the visitor to a site generates one or more "requests" for information from one or more servers (some of these servers may belong to other firms). Each request is recorded as a separate entry. Each page viewed by a visitor to a site may generate multiple requests for files – sometimes a separate request is generated for each icon on the page. The common logs contain the following information, and are useful for generating aggregate site statistics:

- Host name or IP address of the computer making the request to the server
- User name registered on the computer making the request (seldom used)
- User name on the local web site making the request (if the reader logs into a secure area of the web site)
- Time stamp - the date and time of the request
- Request - the actual HTTP request, including the names of the files requested
- Status code - the code for the resulting success or failure of the request
- Transfer volume - the number of bytes sent to the reader's browser as a result of the request.

A major problem in using the common log data for marketing purposes is that there is no continuity to the data recorded by the server. Because the Internet is stateless, each request is transacted between the client and server separately and recorded as a separate entry in the log file. Note also that each request may take a different network route from the client to the server. In general, it is difficult to determine from the common logs the identity of the user, and even

whether the requests are associated with the same user session. To overcome this problem, some servers incorporate more sophisticated approaches[3]:

- The server generates and appends a time-stamped "cookie" (a unique identifier) to the client. The client browser sends the cookie along with each request it sends to the server – this exchange of cookie information between the client and the server helps maintain continuity during a session. The server also appends the cookie ID to its logs, thereby enabling the web site to track the sequence of activities undertaken by a user during a session. Cookies do not typically permit data from different sessions to be linked together in a coherent manner. Cookies are also specific to a server, which means that it is difficult to use them to track users as they move from one web site to another. Note that the user name or registration is not required for the cookie-based approaches to work. All that the server needs is a unique cookie ID that it generates.
- The server appends an ID to the URL that it sends back to the client browser, creating a "virtual page" just for that user. Using this ID, the server then keeps track of requests made by the user during a session, in a manner similar to that of cookies.
- The site requires users to register before they get access (e.g., travelocity.com). This is the most reliable method for tracking customers across visits to a site.

Log file data can be aggregated and organized to facilitate decision making, particularly for improving site design. Aggregate site statistics are widely used by companies because of the availability of software for summarizing the log data. Such statistics as the most viewed pages, server requests by hour of day and day of week, the top referring sites, etc. are easily compiled with these software.

From a marketer's perspective, individual-level data are likely to be far more important than aggregate data. Understanding the behavior of individual customers would allow marketers to customize their offerings to better meet the needs of each customer. Ultimately, this is the major promise and benefit of digital interactivity (Wind and Rangaswamy 1999). Exhibit 2 summarizes the various measures and data that online marketers can potentially use to enhance their understanding of how their online marketing tactics influence customer decision making. It

[3] These methods have generated considerable interest among both marketers and privacy advocates. The privacy issues will be hotly debated in the coming months, especially with regard to the types of information that are collected unobtrusively at a site, and how such information is used.

is clear that log files (suitably decoded and combined with other data sources) could be a rich source of information to drive marketing decisions. However, there are at present a number of limitations of individual-level data obtained from web servers. First, these data only track customer behavior, and hence, they are limited in the extent to which they can provide explanations for observed behavior. Second, without some requirement for registration, it is difficult to obtain longitudinal data that tracks the same customer across time. For example, at auctions sites such as ebay and Yahoo!, the companies can get longitudinal data on the bidding behavior and past purchases made by a member, but they cannot fully track all the information a particular member searched across visits to the site. This is because members are required to register for bidding, but not for browsing the site. Third, it is difficult to track customer visits across competitive sites, except for customers who are members of research panels, such as those of Media Metrix. Companies, such as Doubleclick.com are attempting to generate data based on tracking customers across sites. These companies have developed technologies so that they receive information from affiliated sites each time any of these sites receives a cookie from a user. This allows these companies to track the cumulative number of (banner) ads that a user is exposed to across visits to different sites even over several sessions. However, such data are presently limited to tracking customers only across sites that are affiliated with these companies.

In spite of the rapid developments in tracking web site traffic, we are now only at the early stages of understanding how best to use these types of data for driving marketing decisions, especially at the level of the individual customer. We need more research to find improved methods to link web site data to marketing decisions and outcomes, and to identify new types of data that would be useful to online marketers. For example, it is becoming increasingly important to segment customers based on the "patterns of search" they undertake at web sites -- the sequence in which they view various pages at the site, or how they search for and sort the information they are presented with. From a practical perspective, such segmentation can be very useful to marketers, because it enables them to recognize a potentially adverse search sequence early in the process and "intervene" to offer helpful tips, thereby increasing the odds

that a sale will result from that customer. For example, if the beginning elements of a search sequence indicate that the customer is likely to be confused by the product options, automated agents can selectively provide information only for products likely to be of interest to that customer. In the web environment, it will become critical to do this type of "just-in-time" marketing research.

3. Integrating Marketing Research, Strategy and Planning

In the short term, the new digital technologies discussed in the previous section will make it faster, easier, and cheaper to conduct marketing research studies. Therefore, even small firms will begin using marketing research to augment managerial judgment. For example, a small online firm can get data about customer behavior (from its web site) that it can use for improving its offerings. Larger firms will be able to evaluate more product ideas and marketing programs even with the same marketing research budget. These developments point to the beginning of a long-term transformation of marketing research. Just as survey research methods developed in the thirties revolutionized the theory and practice of marketing research (see, for example, Converse, 1987) the digital medium will have a comparable impact in yet again transforming marketing research. We see the following major trends:

Marketing research will become responsible for on-going data collection and its use. While there is always a need for one-shot studies, marketing research is becoming more closely integrated with on-going decision making. Organizations are investing heavily in information technology infrastructure and communication networks to improve their performance and future prospects. Technologies such as UPC scanners and web server logs, and marketing tools such as customer loyalty programs help firms collect data on a continuous basis at the level of the individual customer. In such environments, marketing research can take on an increasingly important role for facilitating organization-wide decision making (see, for example, Curry 1993;

Hughes 1994). Marketing research is thus becoming a vehicle for tracking customer loyalty, estimating the lifetime value of customers, building customer relationships, developing cross-selling programs, customizing marketing programs, and enhancing customer retention (Berry and Linoff 1997).

Many organizations now have a new function called Marketing Information Systems (MkIS), which is charged with harnessing marketing-related information and distributing and facilitating its use within firms. As marketing research becomes an integral part of running a business, data and insights will be needed on a just-in-time basis for large and diverse sets of people working on different projects. If marketing research data can be deployed in this way, it may even display the "increasing returns to scale" that seems to characterize knowledge assets -- the more the research is used, the more value it generates, and the more secure its role within the firm. Consider, for example, how Wal-Mart uses sophisticated database and decision support tools to help its store managers improve store performance. Its core software contains over 30 million lines of proprietary code, which allow the company to determine the profit (or loss) on each one of 65 million shopping baskets purchased every week at its 3,000 stores in seven countries (Information Week, December 9, 1996). By enabling continuous tracking of the sales of each item at each store, the system allows managers to manage the product mix, expedite price changes, plan promotions, etc. This type of "continuous data-based research" will become more prevalent in more firms. Even one-shot studies can take advantage of this continuous data collection. For example, MCI developed its popular "Friends and Family" program by analyzing the calling patterns of its customers and finding that most households call at most 12 different telephone numbers on a regular basis.

As we look further ahead, technology supporting individual-level data will allow us to do research in fundamentally different ways. To highlight the evolution in marketing research, consider a typical research project being conducted today. A company commissions a study to help identify opportunities in a new market, profile the most attractive customer segment, and design a new product and marketing program to appeal to this group. The steps might include (1) surveying a representative sample from the total population, (2) conducting focus groups and survey research to identify consumers' needs and wants, (3) identifying the largest and most profitable segment, (4) designing and testing alternative marketing programs with the target segment using concept tests, in-home use tests, field experiments and test markets, (5) selecting the program that scores best, and (6) mass-marketing this program to the target segment. Throughout this process, statistical methodologies are used to estimate the likelihood that consumer attitudes and behaviors observed in the sample will generalize to the population.

Using digital technologies, organizations may take a very different approach. They can: (1) gather information on an ongoing basis from a variety of sources on the entire population of target customers, (2) use geographic, demographic, life stage and purchase information to infer customers' needs and wants, (3) measure the value and responsiveness of each individual, (4) tailor marketing programs to unique customer requirements, and (5) continue to track customers' characteristics and behavior over time, making appropriate changes to the marketing mix where necessary.

The newer approaches treat customers as individuals rather than averages, so their unique needs and wants can be better satisfied. Marketing research becomes an integral part of managing customer relationships. Manufacturers, retailers and service companies are then directly involved in analyzing and interpreting customer information, rather than delegating the

task to outside consultants. The data are dynamic and timely, providing continuous feedback on the performance of marketing decisions. The focus is on optimizing the entire set of marketing mix variables, rather than addressing one element at a time. Because the data are collected on the entire population, issues of sampling and statistical estimation become less important. Finally, using individual-level analysis, companies are increasingly able to customize their marketing activities through catalogs, direct advertising and promotions, online shopping sites, customer service, store- and customer-specific products and planograms, information kiosks and electronic shopping assistants.

To fully realize this level of integrated research, organizations will need powerful information systems to collect and analyze individual-level data, and the ability to customize marketing programs. Database marketing companies have already made significant progress integrating public sources of customer data (including census data, birth records, bankruptcy and divorce proceedings, real-estate transactions, motor vehicle registrations, and driver's license information) with private data sources (such as point-of-sale data, credit/debit card and ATM transactions, telephone billing information, membership lists, home shopping activities, electronic ticketing, etc.). See, for example, David Shepard Associates 1995.

Customers will become an integral part of the marketing planning process. As an example, consider how the educational division of Texas Instruments (TI) has used its web site (ti.com/calc) to establish links with potential customers in schools across the country. The company developed the highly successful TI-92 calculator by getting customers closely involved in the product's development. TI posted the proposed specifications of the product at its web site, and offered an online demo simulating its functionality. It then invited feedback from members of various discussion groups where high-school teachers tended to congregate. Based

on the feedback, major changes were made to the product, which led to further dialog with its online community of "marketing research respondents." The continuing dialog with these participants was instrumental in making many enhancements to the product. When the calculator was introduced, the final specs were also put on the web site. This offered a simple way for teachers to download documents to develop proposals to their school boards for purchase of these calculators. When another company came up with a similar design within a few weeks, many of the teachers were offended that the other company "had stolen their design" and boycotted the competitive product. The success of this approach at TI and at other companies (e.g., Netscape) suggest that perhaps organizations might establish a panel of customers/prospects (with suitable incentives) to provide ongoing research inputs. There are, however, some limitations to setting up such panels, including the possibility of potentially leaking useful product ideas to competitors, and unlike traditional panels, the questions/responses may lack standardization.

In a different approach to involving customers in ongoing research, companies such as Dell and Paris Miki's (one of the world's largest eyewear retailer) have used customer inputs to link up research, product development, and production. Dell allows customers to design "their own computer" by selecting product options listed at their web site. It then custom-produces a computer with precisely those options. In this process, the company collects considerable useful information about customers' preferences for various product options. Paris Miki's uses multimedia technology to more closely link customer preferences to its production capabilities. In its retail outlets, a customer can make a digitized image of her face, and try on various virtual frames with different lens shapes, etc. to custom-design a rimless pair of eyeglasses. The selected design is then produced on site using sophisticated production technologies. The

approach used by Dell and Paris Miki's overcomes one of the common problems for which traditional marketing research was developed, namely, not knowing what exactly each customer (or customer segments) will value before the company produces the product. Further, by having information about individual customers that competitors do not have (e.g., facial features of a customer), a company can offer better value to its customers.

There will be increased use of controlled experiments: The digital medium facilitates, and often demands, continuous experimentation. In the digital medium, the real world serves as the laboratory, and offers much higher levels of control at lower costs than is the case with the physical medium. For example, it is a simple matter for online stores to change prices (if required, only for selected consumers) and determine the impact of this change on sales. The digital medium also offers the potential for constructing much richer and more interactive stimuli for experimentation. This is particularly useful for new product testing using such techniques as information acceleration (Urban, Hauser, and Roberts 1990), where future scenarios depicting contexts for product use could be presented more effectively.

New methods and models will be developed for deriving insights from large data sets. Companies are now building data warehouses which bring together information from multiple sources, organized in a common format with consistent definitions for keys and fields, and providing input/output facilities for decision support. Data mining, a field of study at the intersection of statistics and computer science, offers a set of tools for extracting "hidden predictive insights" from large data sets, especially when the structure of the data is unchanging but the data elements themselves are continuously changing – as in scanner data. Data mining tools are now being developed to sift through the marketing data to identify patterns and relationships using traditional analytic tools like regression, cluster analysis, and decision trees,

and newer technologies like Artificial Neural Networks, Genetic Algorithms, and Bayesian Networks. For more details about some of the newer techniques, see Fayyad and Piatesky-Shapiro (1996) and Levin and Zahavi (1999).

The Internet is making possible the collection of even more data. It is just a matter of time before we have "marketing research agents"[4] that scour the Internet to do such activities as gather data from specified web sites at specified times, administer questionnaires (e.g., bizrate.com), or generate top line reports on demand.

In such data-rich environments, deriving useful insights will depend on the ability to manipulate large amounts of data and automate parts of the analyses. Inaccessible or unusable data serve no purpose. Without the deployment of "intelligent data interpretation agents," potentially useful information and insights will lay buried inside vast databases or will be lost in cyberspace. Already, "intelligent agents" on the Internet can help search for specific information (e.g., hotbot.com), compile information according to user preferences (e.g., pointcast.com), or correlate information in useful ways (e.g., firefly, now part of Microsoft). There is today an urgent need for developing data mining techniques and agents that serve the marketing community, and for evaluating the relative merits of alternative approaches in various marketing research applications.

4. Conclusions

In this paper we have highlighted several novel ways in which digital technologies are transforming marketing research and some of the advantages and limitations of marketing research methods that use these technologies. There is very little academic research to guide practitioners in choosing between these methods, or in effectively using a particular method in a

[4] An agent is a an autonomous software entity that carries out specific operations on behalf of an Internet user or another software program, based on specified goals and preferences.

given application. In fact, in the Acorn/IIR surveys reported earlier, practitioners bemoaned the lack of academic research to guide them in using digital marketing research methods. We see an urgent need for research in several areas:

1. Data collection: We need more research that carefully explores the relative efficacy of digital data collection methods for obtaining various types of primary data – awareness, attitudes, motivations, preferences, intentions, emotions, and choices of customers. What is the reliability and validity of these data compared to data obtained through traditional marketing research? What techniques (e.g., sampling procedures) will improve the quality of marketing research currently being conducted online? To address these issues, we need a better understanding of how people behave in digital environments.
2. Data analysis and interpretation: We also need more research to develop robust models and methods for interpreting large amounts of data. Much of the research addressing these issues (e.g., data mining) is currently being done in computer science and statistics. Marketing scientists need to get actively involved in this research stream to adapt existing methods to the needs of marketing managers, and to develop new methods that take advantage of the structure of marketing data and decision problems. In particular, we anticipate a need for methods that combine individual-level behavior information (e.g., what a customer has bought in the past), attitudinal information collected on a sample (either online or offline), and descriptive information about customers (e.g., data such as those supplied by Acxiom and Polk).
3. Data use: Finally, we need more research to understand how the digital environment is transforming managerial roles and responsibilities, and the role that marketing research can play in improving decisions and actions in this environment. Having information is not enough. It is more important to deploy the information when and where it counts. This may require "mass customizing" of data and reports dynamically tailoring them to the needs of individual managers. This can be a complex process in large organizations. Further, managers must first be willing to trust the information produced through marketing research

(Moorman, Zaltman, and Deshpande, 1992). It thus becomes important to understand what aspects of the new research methods promote managerial trust in the data, and the factors that promote the deployment and use of marketing research data and insights at the various points of decision making within an organization.

What actions should managers be taking now? At present, the best approach is to devote at least some of the marketing research budget to experimenting with the new methods. Managers in companies selling primarily digital products (e.g., software, entertainment, music) should take immediate steps to experiment with all the new types of marketing research that we have discussed in this paper. The online medium (i.e., marketspace) is quickly becoming an important, if not the primary channel, in the U.S. for selling these products. As this happens, research using digital technologies may ***supplant*** traditional research as the primary vehicle for obtaining data about customers, competitors, and markets for these products. Managers in companies selling primarily nondigital products (e.g., cars, steel, fast food), should start experimenting with the new research methods selectively, using them in areas where they are likely to be most beneficial. Research based on digital technologies are likely to ***complement,*** not supplant**,** traditional research, at least for the next few years. Senior management in every organization should begin devising ways to deploy marketing research data and insights onto the IT systems and digital backbone that most organizations are putting into place.

In recent years, there has been growing skepticism about the role and value of traditional marketing research, especially quantitative research, in influencing managerial decision making (Wind 1997). Many managers believe that marketing research often does not deliver the information that they really need, that it costs too much for what it actually delivers, and even when it does deliver useful information, it comes in late. The digital medium offers an opportunity to address these limitations of existing research approaches.

Concept Test (20 minute)	**Production/ mailing cost**	**Recruiting cost**	**Respondent incentive**	**Interviewing cost**	**Stimulus complexity**	**Sample representativeness**
Mail concept with printed survey	High	>$0 (list)	$1 + $10	$0	Low	Low (due to low response rate)
Mail concept with telephone interview	Moderate	$10	$10	$25	Low	High
Pre-recruit personal interview	Moderate	$10	$20	$40	Moderate	High
Pre-recruit computerized study	Moderate	$10	$20	$10	High	High
Mall intercept Computerized study	Moderate	$5	$10	$10	High	Unknown (probably moderate)
Web intercept	Low	>$0 (web ads)	$10	$0	Moderate	Unknown (probably low)
Pre-recruit web interview (ID #)	Low	$15 (low incidence)	$10	$0	Moderate	Moderate (computer users)

Exhibit 1: The above table summarizes the costs and benefits of various research methods for conducting concept tests. Some online methods (e.g., Pre-recruit web interviews) cost less than offline methods, but still deliver acceptable levels of stimulus complexity and sample representativeness.

Customer Decision Stages	Measures	Data sources
Awareness and Search	**Total pages delivered**	**Enhanced log file**
	Cumulative number of visits	**Enhanced log file**
	Unique visitors	**Enhanced log file (e.g., with cookies)**
	Visitor profiles	**Registration**
	Aided/Unaided recall	**Online intercepts/panel surveys**
	Clickthroughs (referrals from other sites)	**Log files/data from affiliates**
Interest and Evaluation	**Incoming links, user sites/groups**	**Enhanced log file**
	Visit duration and depth	**Enhanced log file**
	Inter-visit duration	**Enhanced log file + registration**
	Requests for information	**Log file/customer database**
	Leads generated	**Customer database**
	Site search usage	**Enhanced log file**
	Brand attitude and knowledge	**Online intercept surveys, panels**
	E-mail activity	**e-mail server database**
Desire and Trial	**Requests for information**	**Log file/Customer database**
	Downloads	**Log files+registration**
	Simulator usage	**Log files/Activity monitoring**
	Preferences	**Activity monitoring/registration**
	Consideration set inclusion	**Activity monitoring, Online surveys, panels**
	Qualified leads	**Customer Database**
	Participation in promotions	**Registration/Database, surveys**
Action	**Online ordering**	**Log file/Database**
	Coupon redemption	**Log file/Database**
	Cross sell/Up sell	**Enhanced log file/registration**
	Store visits (e.g., competing stores)	**Surveys/Channel partner database**
	Automated replenishment	**Transactions database**
Post-purchase relationship	**Customer satisfaction**	**Surveys/Resource usage at site**
	Repeat purchase intent	**Surveys**
	Repeat purchase rate and amount	**Enhanced log file+ registration**
	FAQ usage	**Enhanced log file/Database**
	Incoming calls	**Customer database+unique ID**
	Share of customer requirements	**Surveys/offline database**

Exhibit 2 summarizes the various stages in an online customer's decision process and the corresponding measures and data that could facilitate our understanding of the decision process. In this Exhibit, we have focused on the decision process at the individual level. The common log formats are best suited for generating aggregate information across customers.

References

Briones, Maricris (1998), "Cheaper Desktops Will Help Net Researchers Corral Clients," *Marketing News*, 32 (23), p.1,17.

Burke, Raymond R., Bari A. Harlam, Barbara Kahn, Leonard M. Lodish (1992), "Comparing Dynamic Consumer Choice in Real and Computer-simulated Environments," *Journal of Consumer Research*, (June), Vol 19. No. 1, p. 71-82.

Burke, Raymond R. (1996), "Virtual Shopping: Breakthrough in Marketing Research," *Harvard Business Review*, 74 (March-April), p. 120-131.

Converse, Jean M. (1987), *Survey Research in the United States*, University of California Press, CA: Berkeley.

Curry, David J. (1993), *The New Marketing Research Systems: How to Use Strategic Database Information for Better Marketing Decisions*, New York: John Wiley & Sons, Inc.

David Shepard Associates, Inc. (1995), *The New Direct Marketing: How to Implement a Profit-Driven Database Marketing Strategy*, New York: McGraw-Hill.

Feder, Barnaby J. (1997), "Test Marketers Use Virtual Shopping to Gauge Potential of Real Products," *The New York Times*, December 22, p.C3.

Fayyad, U.M. and G. Piatesky-Shapiro (1996), *Advances in Knowledge Discovery and Data Mining*, MIT Press.

Hughes, Arthur M. (1994), *Strategic Database Marketing*, Chicago, IL: Irwin Professional Publishing.

Kannan, P.K., Ai-Mei Chang and Andrew B. Whinston (1998), Marketing Information on the I-Way, *Communications of the ACM*, Vol. 41, No. 3, p. 35-43.

Lawrence, Steve and Lee Giles (1999), "Accessibility and Distribution of Information on the Web," *Nature*, Vol. 400, p. 107-109.

Levin, Nissan and Jacob Zahavi (1999), "Data Mining in Digital Marketing," Forthcoming in *Digital Marketing*, Jerry Wind and Vijay Mahajan (Editors).

Malhotra, Naresh K. (1993), *Marketing Research*, Prentice Hall, NJ: Englewood Cliffs.

Malhotra, Naresh K. Mark Peterson and Susan Bardi Kleiser (1999), "Marketing research: A state-of-the art review and directions for the twenty first century," *Journal of the Academy of Marketing Science*, Vol. 27, No. 2, pp. 160-183.

Moorman, Christine, Gerald Zaltman, and Rohit Deshpande (1992), "Relationships Between Providers and Users of Marketing Research," *Journal of Marketing Research*, Vol. 29, No. 3, p. 314-328.

Reibstein, David J. (1999), "Who is Buying on the Internet, 1999?," Forthcoming in *Digital Marketing*, Jerry Wind and Vijay Mahajan (Editors).

Rickard, Leah (1993), "Goodyear Test-drives the Visionary Shopper for Marketing Research," *Advertising Age*, Vol. 64, No. 45 (October 25), p. 24.

Urban, Glen L., John R. Hauser, and John H. Roberts (1990), "Prelaunch Forecasting of New Automobiles," *Management Science*, Vol. 36, No. 4, (April), pp. 401-421.

Weissman, Rachel X. (1998), "Online or Off Target?" *American Demographics*, Vol. 20, No. 11 (November), p. 20-21.

Wind, Jerry (1997), "Start Your Engines," *Marketing Research*, Vol. 9, No. 4, p. 4-11.

Wind, Jerry and Arvind Rangaswamy (1999), "Customerization: The Second Revolution in Mass Customization," Working paper, The Wharton School, University of Pennsylvania, PA 19104.

Yoffie, Amy (1998), "The 'Sampling Dilemma' Is No Different On-line," *Marketing News*, Vol. 32, No. 8, (April 13), p. 16.