



Leverage Your Web Channel: The Keynote Methodology for Gathering Insights into Online Customer Behavior

Keynote Systems, Inc.
777 Mariners Island Blvd.
San Mateo, CA 94404
Main Tel: 1-800-KEYNOTE
Main Fax: 650-403-5500
product-info@keynote.com

Table of Contents

Executive Summary.....	3
Why Is Customer Experience Management Critical?.....	3
Measuring Customer Experience.....	4
A “Blueprint” of the Customer Experience.....	5
Close-Up and Wide-Angle Views of the Customer Experience.....	5
Customer Experience Evaluation Produces Actionable Insights.....	6
The Keynote Approach to Measuring Customer Experience.....	6
Real People Interacting with Real Web Sites.....	7
Intent-Based Context and Scenario-Based Testing.....	7
Site Evaluations Conducted in Natural Settings.....	8
Research Expertise.....	8
Keynote Research: Combining the Best of the Research Models.....	8
Traditional Market Research and Usability Methods.....	8
Specialized Research Tools and Techniques.....	10
A Comprehensive View of Your Online Customer.....	12
Keynote Research Applications and Usage.....	12
Market Discovery—Assessing Market Needs to Determine Strategic Direction.....	12
Exploration—Deeper Understanding of Needs and Possible Solutions.....	13
Evaluation—Making Sure a Proposed Solution Will Be Effective.....	14
Tracking—Metrics for Ongoing Management of Web Effectiveness.....	14
Methodological Issues and Concerns.....	14
User Objectives and Questions.....	14
Keynote Research Panel.....	15
Samples for Specific Evaluations.....	15
Sample Size and Statistical Significance.....	15
Response Rate.....	16
Panelist Variables.....	16
Panelist Incentives.....	16
Fraud Checks.....	16
Scenario-Based Testing.....	16
Comparing Multiple Web Sites.....	17
Within-Subjects Designs.....	17
Learning Effects and Counterbalancing.....	17
Sample Size for Within-Subjects Designs.....	17
Previous Exposure Effects.....	17
Between-Subjects Designs.....	17
Cross-Sectional and Longitudinal Designs.....	18
The Keynote Evaluation Development Team.....	18
Keynote Consultants.....	18
Keynote Research Team.....	18
Summary of Keynote Methodology.....	19
Appendix A: Keynote Checklist for Formulating Studies.....	20

Executive Summary

Online businesses that don't provide an engaging, hassle-free customer experience can't convert browsers into buyers—or first-time buyers into repeat customers. The first step toward improving a Web site's conversion and customer retention rates is to understand what customers actually experience on the site and which elements most impact overall satisfaction. But it is impossible to predict customer experience without reliable, interpretable data from real people interacting with the site.

Keynote's research methodology is based upon an approach to quantifying and analyzing an organization's online presence, called Customer Experience Management. This methodology provides a window into the total Web experience from the customer's point of view. Keynote customer experience evaluations reveal not only what customers do online, but also why they do what they do.

Keynote's in-depth evaluations can provide key insights into a wide variety of marketing concerns, including:

- Web Site Functionality
 - Making sure users can accomplish crucial tasks easily and accurately
 - Pinpointing where and why users have trouble on the site
- Customer Conversion and Retention
 - Converting browsers to buyers
 - Exceeding user expectations and maximizing satisfaction and loyalty
- Competitive Positioning
 - Comparing the experience on the site to that of other sites
 - Differentiating the site from the competition

- Brand Impact
 - Conveying the site's unique value proposition effectively
 - Ensuring consistency between online and offline branding strategies
- Relationship Marketing
 - Appealing to specific types of users
 - Identifying unique needs among groups of users

This paper explains in detail what the Keynote Methodology is, how it differs from other online research offerings, and how specific methodological concerns are addressed.

Why Is Customer Experience Management Critical?

Online customers can wield their purchase power with a simple mouse click, effortlessly jumping from one online business to another. Given such low consumer switching costs, the burden is on Web marketers not only to attract site visitors, but also to retain them by offering compelling and productive online experiences. Enterprises that fail to do so will be unable to convert browsers into buyers and retain loyalty from these customers over time. Even when the goal is not an online purchase, customer experience is a decisive factor in whether people will use the website to research information, answer customer support questions, or find out more about employment opportunities or investor relations information.

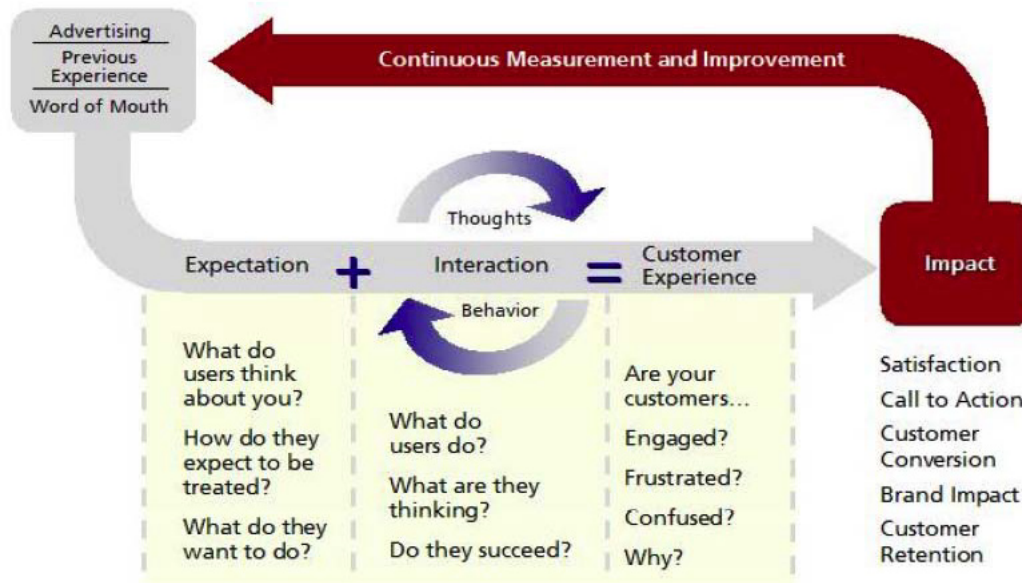


Figure 1. The Keynote methodology captures the complete customer experience. Customers' expectations and interactions drive the customer experience, which in turn, drive the impact and ongoing site measurement and improvement.

In order to improve their online initiatives, progressive companies are making an attempt to track customer behavior and evaluate the end-user experience of their site. But these companies are quickly realizing that clickthrough data and log file metrics are not providing the information required to make strategic business decisions. As the focus for research shifts from site information to customer information, better methods are required to understand the customer's point of view. The first step in addressing a Web site's conversion and customer retention rates is to measure what the customer actually experiences on the site and which elements most impact overall satisfaction.

Measuring Customer Experience

Customer experience on the Web is difficult to predict without reliable, interpretable data. Customer experience is composed of complex interactions between the Web site itself and the thoughts, feelings, behaviors, habits, expectations, and social references that the customer brings to the situation!¹ It is not the objective reality of the Web site that needs to be analyzed, but the *subjective* reality of the customer—the customer's perception and interpretation of the site. To add to this complexity, each individual visiting a site has his or her personal history, creating many different subjective realities for the Web marketer or researcher to understand.

¹ For a good discussion of the "customer's experience" see Schmitt, B. H. (1999). *Experiential Marketing: How to get customer to sense, feel, think, act and relate to your company and brands*. New York: Simon & Schuster. For a review of the degree to which emotions affect choices and behavior, see Goldman, D. (1995). *Emotional Intelligence*. New York: Bantam.

Analyzing non-obtrusive observational data, such as sales figures and server logs, cannot uncover such information. For example, with server log data it is difficult to determine whether a customer is lingering on a site because of interest or confusion. The most reliable method to capture the actual customer experience is to have many customers try the live site, while simultaneously gathering both behavioral and subjective data from each individual.

A “Blueprint” of the Customer Experience

Keynote evaluation methodology is designed to provide a window into the total customer experience, reducing uncertainty for decision makers. The Keynote approach combines the best aspects of market research and usability testing techniques in order to build a blueprint of the customer’s point of view, from brand awareness to ease-of-use to overall satisfaction.

This blueprint maps directly to the features, functions, and messages of the site itself, providing insights into how specific elements impact the overall quality of the experience and users’ subsequent likelihood to return to the site. Thus, marketing executives discover how they should allocate resources for maximum impact, while designers gain insight into why particular features and functions are either working or not working as planned and how to best modify them.

Close-Up and Wide-Angle Views of the Customer Experience

Keynote’s technology is intentionally flexible to enable clients to address a variety of research questions. For an in-depth perspective, Keynote evaluations can provide a deep understanding of specific aspects of the customer experience, such as what URLs individuals followed and why. Keynote evaluations can also reveal a broader picture of the customer experience, such as how the overall experience changed after a site redesign.

Other uses of Keynote’s technology include: examining what sites individuals visit when asked to research and buy a particular item; comparing the customer experience on different sites to pinpoint competitors’ strengths and weaknesses; assessing design changes by evaluating the site before and after changes; and evaluating hypotheses with true experiments to develop causal theories for customer experience outcomes.

Data to Inform Decisions

Keynote evaluations provide clients with data to inform a variety of business and design decisions. Below are some of the common concerns that Keynote’s methodology and technology can address:

Business Strategy Issues

- Do users understand the site’s value proposition? Does their perception of the value proposition change after site usage?
- How does user experience on the site compare with competitors’ sites?
- After interacting with the site, are users likely to come back? Why or why not?
- Is the actual customer experience consistent with brand positioning? Is it consistent with the offline brand?
- What features are users expecting to see on the site?
- Are particular types of users (e.g., novice users, power users) reacting differently to the site? What special needs do particular groups have?

Design Issues

- Can users accomplish critical tasks, such as searching and registering? If not, why?
- What paths do users take in accomplishing critical tasks? What dead-ends do they encounter? Where do they get lost?
- At what point in the process of pursuing specific tasks do users fail or give up? Why?
- Do users notice and make use of particular features on the site?
- How much time and effort does it take to accomplish critical tasks? How can this best be reduced?
- Do users read and make use of information provided? Do users have enough information?

Customer Experience Evaluation Produces Actionable Insights

Site improvements from traditional usability testing are well documented. Reported improvements range from 75 percent to over 200 percent in the usability metrics of a site.² Keynote evaluations expand upon this method by providing strategic information not found in traditional usability testing, without additional costs. The Keynote approach evaluates Web usability issues within the larger perspective of brand positioning, competitive intelligence, likelihood of return visits, and overall user satisfaction. This type of market research information is critical in reducing uncertainty and avoiding losses that might result from poor decision making.³

The Keynote Approach to Measuring Customer Experience

Keynote evaluates customer experience by inviting large samples of individuals to interact with a live Web site. Panelists log on where they normally access the Web and use a small downloadable companion to Microsoft's Internet Explorer, called the Keynote Connector. Users then pursue a predefined set of tasks, such as registering or using a site's shopping cart feature. Keynote's technology records user behavior (e.g., the URLs they follow, time spent on each page, and number of page views) and provides question prompts and opportunities for making open-ended comments.

Users typically begin by answering background, expectation, usage, awareness, and brand positioning questions before interacting with a site; and end with satisfaction, future intentions, and a re-measurement of brand positioning questions. Keynote automatically compiles this data and presents it via a Web-based interface for easy analysis of top-level concerns. The software provides the opportunity to drill down from behavioral data to verbatim comments from users, making the data easily interpretable to inform decisions.

² Nielsen, J. (2000). *Designing Web Usability: The Practice of Simplicity*. US: New Riders Publishing. See also, Spool, J., et al (1998). *Web Site Usability: A Designer's Guide*. US: Morgan Kaufman Publishers.

³ Duboff, R. & Spaeth, J. (2000). *Why Market Research Matters: Tools and Techniques for Aligning Your Business*. New York: Wiley.

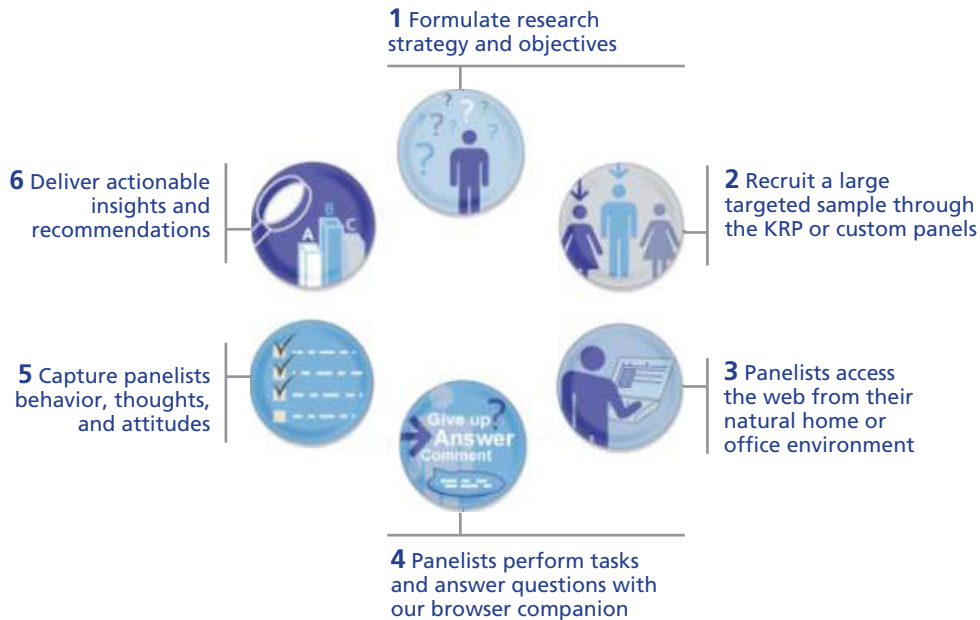


Figure 2. How Keynote Works. A Keynote site evaluation begins by determining the objectives, and then inviting panelists to log on to the Internet in their normal settings using a browser companion that provides instructions and prompts questions to capture their behavior, spontaneous comments and answers to questions.

Real People Interacting with Real Web Sites

Keynote evaluates customer experience by collecting detailed qualitative and quantitative data from large samples of individuals (typically 200 to 800) as they attempt a series of real-life tasks on the Web. Keynote samples users according to target customer profiles from the Keynote Research Panel (VRP) of more than 160,000 Web users, or directly from a client's private panel of actual customers.

In order to capture verbatim comments from users accessing the site, Keynote studies intercept live users and ask them to instantly participate in Web site evaluations. This results in the ability to associate real customer comments and link them to quantitative behavioral data (e.g., clickstream statistics or page views). Capturing the experience of real Web users is the only way to obtain insight into users' subjective thoughts and feelings about a site.

Intent-Based Context and Scenario-Based Testing

Without knowing what customers are trying to achieve, it is impossible to know whether or not they have been successful. In Keynote evaluations, the user's intentions are known. Using the Keynote Connector, a small download that works with Microsoft Internet Explorer, the user pursues a predefined set of tasks (such as registering or using a shopping cart) in a method known as scenario-based testing.

With scenario-based testing, users' goals and intentions are clearly understood and success rates can be determined and compared across the spectrum of users.

Site Evaluations Conducted in Natural Settings

Because Keynote's technology enables remote site evaluations, panelists can participate from any location, at any time of day. Participants access the computer they use every day in their homes or offices, from their own Internet connections, browsers, and computers without having to conform to the constraints of a more artificial testing environment.

By testing in a natural setting, the evaluation experience more accurately represents users' normal Web use conditions. It provides more accurate study results by minimizing interviewer or moderator bias that can arise in a lab or focus group settings. In addition, the flexibility to participate from a variety of locations, such as home, work, or school, at points around the globe, increases the spectrum of potential study participants yielding more accurate test results. Remote settings also offer anonymity, encouraging panelists to express thoughts and feelings with candor.

Research Expertise

A team of professionals with backgrounds in experimental psychology, market research, quantitative management consulting, mathematical modeling, Web usability, and technology publishing develop and refine the innovative Keynote approach. The approach has been applied to a wide variety of Web sites, producing results consistent with other sources of customer experience data while providing more actionable insights.

The logic and benefits of the Keynote approach can be best understood in comparison to current alternative methods, described in the next section.

Keynote Research: Combining the Best of the Research Models

Keynote takes a unique approach to evaluating customer experience, leveraging the most critical elements of traditional market research and usability methods, and also combining both behavior and attitudes to gain unique insight. Making compromises on your audience size or the data collected is no longer required. The Keynote approach combines the advantages of multiple research methods while still providing excellent predictive insights. Contrasting Keynote with some of these standard, traditional research methods highlights Keynote's innovative approach.

Traditional Market Research and Usability Methods

Some of the most commonly used research methodologies include behavior tracking and log files; focus groups; usability studies; consultant reviews; online surveys; and robotic agents. Each methodology will be reviewed below, with information on how Keynote adds unique value and expertise to these kinds of studies.

Behavior-Tracking Tools and Log Files (Traffic Analyses)

Behavior-tracking tools, (e.g., Web server logs) may reveal what users did on a site, but not why they did it. Analysis of log file data alone does not reveal whether a user is abandoning a full shopping cart because he changed his mind after seeing the shipping prices, or whether he did not feel comfortable giving his credit card number over the Internet. Without knowing a user's goals, it is impossible to interpret whether or not he or she is successfully achieving them.

Keynote's Click Stream and "Gold Path" Analyses address the need to understand user intent by employing a scenario-based testing approach. This method, common in traditional usability lab settings, establishes a uniform set of goals (called objectives) that all users pursue. Because user intent is a known variable, Keynote can operationally define and measure success rates for particular tasks. The results can then be linked to qualitative comments and user satisfaction ratings. It is also possible to design open-ended objectives to allow for user-driven exploration of the site.

Focus Groups

Traditional focus groups can provide qualitative data and insights throughout the design process. Yet they can test only small samples, providing insufficient data points for making critical business decisions. In addition, qualitative data depends on interpretations by the moderator and may result from idiosyncratic test participants. Raw data is usually in the form of session videotapes that are difficult to analyze beyond the moderator's report.

Keynote Concept Testing can employ large samples (typically from 200 to 800 individuals) in distant geographic locales. Panelists evaluate sites using their own computer equipment, in their natural settings. Both quantitative and qualitative data is easily accessible for further analysis via online reports.

Usability Tests

Usability tests typically require users to test a site in an unfamiliar setting, often with unrealistically superior computer equipment – conditions that may lead to inaccurate assessments of a site's performance. These tests also require highly skilled moderators, special labs, and complicated logistics for participants, making them expensive and time-consuming. A typical usability study can cost \$40,000 and take as long as six weeks.

Keynote Concept Testing is also useful for quantifying reactions and providing adoption rates. Concept Testing involves getting the target market to learn about the product offering idea and estimate whether they would adopt it. It basically provides insurance for the company that the product will be successful, or can provide an effective way to kill a concept that the company has doubts about.

Consultant Reviews (Heuristic Evaluations)

In this method, experts review Web sites against a checklist of normative standards and general best practices. While this method may uncover problems and act as a starting point in assessing design issues, its effectiveness depends on the level of expertise of the evaluator and the quality of the information on which the checklist was constructed. Keynote Research Best Practices Briefs (Professional Services recommendations based on hundreds of evaluations). In contrast to the normative approach just described, Keynote provides empirical evidence. Consultants or clients can use Keynote evaluations to test hypotheses derived from a normative analysis and to substantiate expert opinions with actual data. This collaboration is especially fruitful with regard to testing and corroborating best practices.

Online Surveys

Surveys poll users via traditional market research studies delivered online or via e-mail. Surveys gather feedback and opinions from large samples of random users by asking a series of closed- and open-ended questions. Principal shortcomings of surveys include their inability to capture behavioral information, making it difficult to assess how people actually interact with a site and whether users complete tasks successfully. Self-reported behavior is an unreliable measure of success. One cannot rely on surveys to identify problem spots on a Web site.

Keynote Private Panel Surveys and Keynote Research Panel Surveys capture extensive behavioral data while users interact with a site, including all pages visited, click streams, time spent on tasks, task success rates, and spontaneous qualitative feedback.

Robotic Agents

Products that use robotic agents are able to assess a Web site in relation to predefined goals, but can only measure the mechanical aspects of the site, and are accurate only to the extent that the robot faithfully simulates human behavior. These robotic inspections cannot assess the subjective aspects of a site or provide insight into a customer's perspective, preferences, satisfaction, or comprehension – all critical elements needed to understand the entire customer experience.

Keynote Evaluations use real people and get real customer attitudes. They collect subjective and qualitative data, along with behavioral data, all of which are important for a complete understanding of the customer experience.

Specialized Research Tools and Techniques

Some of the more specialized research methodologies include competitive testing; syndicated research rankings; pop-up site surveys; field testing; and card sorting techniques.

Competitive Testing

Competitive testing can reveal strengths and weaknesses of an online presence against key competitors, providing strategic direction on how to best compete. But competitive testing is very difficult to do well with traditional methods.

Keynote Competitive Testing provides controlled, experimental ways to compare competing products through within-subjects and between-subjects designs.

Syndicated Research Rankings

Syndicated research reports (e.g., Netratings/JD Powers/Bizrate) can be helpful in showing basic site traffic metrics and can be a good way to get strategic direction for an industry, as they can incorporate large samples, more sophisticated analytics, and 3rd-party objectivity. Unfortunately, many traditional approaches use surveys that do not capture sufficient depth to provide real customer insights into how companies can improve most effectively. Worse, many rankings are not based on customers reactions at all, but on analyst opinions, which can be misleading. Keynote currently offers two types of syndicated research products.

Keynote CE Rankings

Keynote CE Rankings deliver a standardized comparison of customer experience across the top sites in an industry. For each study, customers are randomly assigned to sites and each sample performs the same tasks and answers the same questions, allowing direct comparison across the sites. The studies provide a competitive benchmark from which clients can evaluate their site's performance against industry norms over time; examine industry-wide drivers of acquisition, loyalty, brand and online adoption; identify competitors' strengths and weaknesses; and learn best and worst Web practices.

Keynote Customer Discovery and Tracking Studies

Keynote Customer Discovery and Tracking Studies evaluate the customer experience while performing tasks related to a specific industry. Customers start at a blank Web page and are asked to use any resource on the Internet to perform a task, such as plan a vacation or find information on a medical condition. These studies provide insight into research or shopping processes across the Web – where customers go, the reasons they go there, and the feelings customers have as they go from site to site. Companies can learn how to better capture the attention of Web users and improve their messaging and site customer experience to increase the likelihood of acquisition and retention.

Pop-Up Site Surveys

Site pop-ups are used in order to understand more about users' attitudes, perceptions, and experiences. They gather qualitative and quantitative data. Common uses include customer satisfaction rankings, user requirements analysis, brand awareness, concept testing, and market analyses. Keynote Intercept Evaluations can be used as pop-up surveys, but also provide the ability to capture more in-depth customer feedback and behavior. The evaluations are designed to download very quickly.

Keynote can provide help with all of these different types of research by incorporating the critical aspects of each approach. It provides the intent-based context (scenario approach) of traditional usability testing; large samples associated with surveys and traffic analysis tools; the qualitative data and verbatim reactions of focus groups and usability labs; the behavioral analysis (click streams, page views, time intervals) of usability labs and traffic analyses; and a realistic setting (with users' normal internet connections) that are only associated with traffic analyses. In addition, Keynote's analysis tools make it easy to see how the quantitative and qualitative data relate to each other and come together to form rich customer insights.

Field Tests

Field tests are useful for obtaining a complete understanding of when, where, and how people interact with a company's products or services. However, observing actual customer behavior can be difficult to implement due to the fact that Web users are scattered across the globe. It is important to be able to understand and measure how a wide spectrum of users navigate and make choices when using the Web.

Keynote Open-Web Research Evaluations are useful for capturing quantified behavior for specific Web surfing goals (e.g., "You're wondering if you can get a better deal on a loan somewhere else. Use the Web to determine if that is true.") Keynote evaluations provide insight into actual customer behavior in real-world settings.

Card Sorting

Card sorting allows companies to understand how customers categorize products, rather than how companies view them, and what nomenclature customers can relate to. Card sorting techniques are useful but difficult to do.

Keynote Nomenclature Testing is good for evaluating effectiveness after initial architecture is completed and for refining link labels. Keynote can also use online card sorting techniques within an evaluation context to capture customers' natural categories and understand how they should be integrated into the site's nomenclature.

A Comprehensive View of Your Online Customer

Table One summarizes Keynote's capabilities compared with traditional approaches that offer just a slice of information. Keynote's methodology incorporates the critical aspects of each approach. It provides the task-based approach of traditional usability testing; the large samples associated with surveys and traffic analysis tools; the qualitative data and verbatim reactions of focus groups, usability labs and surveys; the behavioral analysis (clickstreams, page views, time intervals) of usability labs and traffic analyses; and a realistic setting (with users' normal internet connections) that are only associated with surveys and traffic analyses. In addition, Keynote's analysis and reporting tools make it easy to see how the quantitative and qualitative data relate to each other and come together to form rich customer insights.

	Task-Based Approach	Large Samples	Qualitative Data	Behavioral Data	Natural Setting	Analysis & Reporting Tools
Focus Groups			X			
Usability Labs	X		X	X		
Surveys		X	X		X	X
Traffic Analysis		X		X	X	X
Robotic Agents						X
Keynote Method	X	X	X	X	X	X

Table 1. Keynote methodology incorporates the critical aspects of each research method.

Keynote Research Applications and Usage

There are many different methods that can be used to conduct research, but it is essential to choose the right tools to accomplish specific research objectives. In order to help clients select the best research methods, Keynote outlines a Research Roadmap following the product development cycle to help clients select from the different tools and methodologies available to marketers.

Stages of the product development cycle can be divided into four general categories: discovery; exploration; evaluation; and tracking. The Keynote applications appropriate for each stage are discussed below.

Market Discovery—Assessing Market Needs to Determine Strategic Direction

Before any marketing endeavor begins, successful ventures will start with deeply understanding the market in order to develop a winning strategy. Discovery and market strategy include the following processes:

- Setting specific business objectives
- Identifying and assembling core business metrics (sales, leads, traffic or support center data)
- Conducting competitive assessments/benchmarking (baseline)
- Determining drivers of customer behavior/customer insight
- Determining the most important segments

The following Keynote applications are appropriate for these objectives:

- **Keynote Syndicated Research.** Understanding competitive strengths, weaknesses, and market needs (see the Keynote Syndicated Research White Paper for more details)

Keynote Syndicated Research compares top industry sites directly and shows how people are currently using the Web for a particular product category. These studies provide comprehensive coverage of an industry, sophisticated analytics, drivers of customer behavior, and best practices. Syndication, with multiple buyers, insures greater value and third-party objectivity.

- **Keynote Competitive Testing.** Learning the strengths and weaknesses of competitors in an apples-to-apples comparison.

Keynote Competitive Testing provides controlled experimental ways to compare competing Web businesses and products through within-subjects and between-subjects designs.

- **Keynote Web Customer Discovery Studies.** Understanding how people currently use the Web in general, such as researching a product or finding a service.

With Keynote's Open Web Research methodology, panelists are taken to a blank page on the Web and asked to spend some time conducting research relevant to a particular goal. They are encouraged to select sites from their Favorites, directly type in the URLs of sites they know, or go to search engines they commonly use.

- **Keynote Intercept Evaluations.** Understanding current users' backgrounds, intentions, success rates and frustrations.

Keynote Intercept Evaluations, with its very fast download, captures in-depth customer feedback and behavior on current Web properties, providing insight into why current customers come to the site, what they do, and why they leave.

- **Keynote Private Panel Surveys and Keynote Research Panel Surveys.** Learning about customer's attitudes, awareness, usage and needs.

Keynote can conduct surveys on private panels or Keynote Research Panels and match survey data sets to evaluation data sets on the same people.

- **Keynote Customized Benchmarks.** Setting up a baseline to compare success of changes longitudinally.

Keynote benchmarks help reveal user experience metrics to complement site metrics.

- **Keynote Click Stream and "Gold Path" Analyses.** Making sense of users current behavior patterns.

Keynote Golden Path analyses reveal the most successful paths through the site.

Exploration—Deeper Understanding of Needs and Possible Solutions

In the exploration phase, businesses need to gather a deeper understanding of particular needs and test out a variety of possible solutions. User interactions and reaction are observed, and unanticipated insights, obstacles and opportunities are assessed. Solutions include:

- **Keynote Nomenclature Testing.** Explores alternatives for categories and labels.
- **Keynote Nomenclature Testing.** Determines how users understand information architecture and product categories for creating more effective Web site navigation.
- **Keynote Concept Testing.** Explores market receptivity of new features or services.
- **Keynote Concept Testing.** Quantifies reactions and provides adoption rates. Concept Testing involves getting the target market to learn about the product offering idea and estimate whether they would adopt it. It provides insurance for the company that the product will be successful, or can provide an effective way to kill a concept that the company has doubts about.
- **Keynote Early-Stage Prototype Testing.** Tests possible solutions in wire frame form, or solutions implemented by other companies, before investing in further development.
- **Keynote Early-Stage Prototype Testing.** Allows companies to get user reactions to a variety of possible solutions to further refine marketing direction. Keynote can test very rough prototypes or wire frames, or can use another company's implementations of a particular feature or solution as a prototype.
- **Keynote Research Best Practices Briefs.** Understanding design principles and industry best practices to help formulate a better solution
- **Keynote Professional Services.** Mines aggregated data based on hundreds of evaluations to produce research and best practices briefs. Keynote consulting and research prove that heuristic evaluations do not replace real customer feedback.

Evaluation—Making Sure a Proposed Solution Will Be Effective

With evaluation studies, user responses are predicted to implemented the solution and prove or refine the product before launch to ensure success. Specific solutions include Customer Experience Testing and Keynote A/B Testing.

Keynote Customer Experience Testing

Keynote can test whether the overall product meets users' needs and preferences and tease apart usability issues from marketing issues, such as pricing and positioning.

Keynote A/B Testing

A/B testing refers to experimental field testing of marketing media or products to determine experimentally which drives better outcomes. This form of testing is especially powerful in comparing new designs to old ones before they are rolled out to larger audiences, optimizing current designs through testing incremental changes, and proving ROI for designs or offers overall.

Tracking—Metrics for Ongoing Management of Web Effectiveness

Businesses cannot manage what they don't measure, and Web site effectiveness is no exception. Tracking research includes customer satisfaction surveys, competitive assessments and ongoing benchmarking of how well the site delivers on key tasks and outcomes. Keynote research can be used for benchmarking and competitive tracking through syndicated studies.

Keynote Benchmarking

Keynote benchmarking studies are used to develop standardized site effectiveness metrics and track progress to goals. When maintaining a family of branded Web sites within a competitive space, it is important to establish a standardized benchmarking system for measuring and monitoring the customer experience across those sites as well as their competitors. By deploying the Keynote Benchmarking Solution, enterprises will be able to:

- Monitor and compare the customer experience across the participant's site
- Compare sites with key competitors' sites
- Understand drivers of customer behavior
- Prioritize development and marketing efforts
- Monitor progress over time using objective metrics

Keynote Syndicated Research Tracking Studies

Understanding competitive strengths, weaknesses and market needs over time is what the Keynote Syndication studies track. (See Keynote Syndication White Paper for more details.) Keynote syndicated research compares top industry sites directly and shows how people are currently using the Web for a particular product category. The studies provide comprehensive coverage of an industry, sophisticated analytics, drivers of customer behavior, and best practices. Syndication, with multiple buyers, insures greater value and 3rd party objectivity.

Methodological Issues and Concerns

The success of the Keynote approach depends on the quality of the questions, the sampling procedures, and the validity of the evaluation methods. The following sections detail specific elements of the Keynote approach and strategies for overcoming common methodological concerns.

User Objectives and Questions

Keynote maintains a library of evaluation scripts with particular objectives and supporting questions that are common to most Web sites (e.g., registration, searching, understanding the core value proposition of the business, or estimating user satisfaction). The Keynote research team conducts internal analyses to determine which questions are most likely to provide insights into the customer experience. In this way, clients gain the benefit from Keynote's collective experience with previous evaluations.

Keynote Research Panel

Keynote has its own growing research panel, the Keynote Research Panel (VRP), which currently contains over 160,000 Web users with thousands of individuals joining each month. The panel is managed to provide the full spectrum of Internet users and to meet high-demand categories, such as high-income shoppers or business decision-makers.

To combat sampling biases associated with self-selection, Keynote recruits panelists through a broad range of sources including word-of-mouth, e-marketing campaigns, affiliate programs, and invitations on portal sites. Keynote also proactively recruits under-represented groups. The Keynote research team runs internal evaluations periodically to ensure that data from the VRP are similar to data that would be obtained from a true probability sample.

The possibility of unknown biases associated with non-random samples is typically of greater concern to the scientific purist than to the business decision-maker, but there may be some practical implications. For example, panelists who join the VRP are likely to be more Web-savvy than the general population and be more open-minded to exploring new Web businesses. As is true of all market research sampling, sample results must be interpreted with the caveat that statistical estimation assumes a true random sample from the population of interest.

Samples for Specific Evaluations

Keynote can construct samples from the VRP to model the on-line population or particular target markets by using quotas on particular attributes. For example, if the Web population were 56 percent men and 44 percent women, a 200-person sample would contain quota targets of 112 men and 88 women. Keynote would then randomly sample men and women from the VRP until these quotas are filled.

Although most clients utilize the VRP for their sampling needs, clients may also recruit participants directly from their own Web sites, customer lists, or RDD (random digit dialing) samples. In all cases, Keynote works with each client to generate a sample that approximates the client's target customer profile.

Sample Size and Statistical Significance

Keynote employs larger samples (typically from 200 to 800 individuals) than traditional usability tests, which tend to use five to eight people. Larger sample sizes provide better insights into brand positioning and a more comprehensive picture of customer experience that businesses require to make strategic decisions. These large samples comprise a variety of users, ensuring the representation of many perspectives and the ability to estimate the magnitude of problems by the percentage of users who encounter them.

Large samples can be analyzed with statistical tests to more accurately interpret comparisons. Larger samples make statistical tests more sensitive for detecting possible differences among groups, or between observed and expected results. Typically, with a sample size of 200, statistics such as chi-squared, t-tests and regression will detect statistically significant differences with a confidence level of 95 percent. In addition, a sample size of 200 is large enough for a statistically meaningful analysis of sub-samples (e.g., types of Web users).

What does "statistically different" mean? Sometimes differences arise simply because of sampling errors, rather than true differences between populations. A confidence level of 95 percent means that it is highly unlikely that a particular difference between distributions arose from sampling error alone, rather than true underlying differences between the distributions.

Response Rate

All market research surveys, usability tests, and customer ratings are affected by response rates, that is, the percentage of invited members of the tester pool who respond within the specified data collection period. Response rate is important, because those who respond may be different from those who did not, rendering the sample less representative. For example, those who choose to participate in an evaluation may be people who are more enthusiastic than usual, and thus the results from this group may not fully represent the entire population. Response rates for Keynote evaluations are similar to industry averages for email-related surveys. Response rates are monitored and regulated to ensure that impact on the representativeness of the sample is minimized.

Panelist Variables

Keynote collects demographic data from each member of the VRP at the time of registration. Keynote then surveys panelists several times per year to update their profiles and obtain more specific data on their purchasing behaviors and interests. Panelists must be at least 13 years of age to participate (18 if the panelist resides in the European Union). In obtaining, storing, and sharing panelist data, Keynote complies with guidelines established by the World Association of Opinion and Marketing Research Professionals (ESOMAR) and TRUSTe.⁴

Panelist Incentives

Keynote offers panelists a token of appreciation for completing an evaluation, such as a gift certificate or opportunity to donate to charity. Keynote panelists sometimes receive incentives to participate in periodic surveys, such as an entry into a cash sweepstakes. However, many Keynote panelists report that curiosity and a desire to improve customer

experience motivate them more than the evaluation incentives do. To discourage “professional” panelists, Keynote prevents individuals from participating in an evaluation more than once a month and more than eight times in a 12-month period.

Fraud Checks

Keynote screens all panelist registrations for possible fraud before admission to the VRP, eliminating panelists who sign up more than once under different names or provide obviously false information. The completion rate is the percentage of invited respondents who complete the evaluation and make a good faith effort to complete the objectives. Keynote routinely excludes panelists that do not meet this “reasonable effort” criterion from the final sample. Of those who respond, the percentage of panelists completing the evaluation with usable data is very high, minimally impacting representativeness of the sample.

Scenario-Based Testing

Like traditional usability tests, Keynote employs scenario-based techniques, where panelists are asked to pursue a structured set of objectives. This approach has the advantage of making both aggregate and individual behavior interpretable. Because users who are provided goals may behave differently from users who have their own goals, Keynote stresses using scenarios and objectives that are similar to what a real customer would encounter. Findings from Keynote evaluations employing scenario-based procedures are highly consistent with findings from log analyses of actual customers, providing confidence that scenario-based testing biases are minimal.

⁴ For the full text of these documents, see http://www.esomar.nl/codes_and_guidelines.html, http://www.truste.org/webpublishers/pub_resourceguide.html.

Comparing Multiple Web Sites

Comparing multiple Web sites presents special methodological concerns. The following sections explain how Keynote addresses these issues.

Within-Subjects Designs

In Keynote's Comparative Evaluation solution, panelists evaluate more than one Web site and attempt the same set of tasks on each site. Each participant thus serves as his or her own control, completing the same objectives on both the client site and competitors' sites (this is called a "within-subjects" design, as opposed to a "between-subjects" design, where each site is evaluated with a different group of panelists). This within-subjects design allows for a direct comparison of users' experiences on the two sites and provides more power for statistical comparisons. The different sites provide an immediate comparison for all statistics. For example, a Web site may have a registration failure rate of 20 percent, which might seem adequate by industry standards. However, a Comparative Evaluation may reveal that the same group of people showed a failure rate of only two percent on a competing site, revealing an important need for improvement.

Learning Effects and Counterbalancing

Keynote's Comparative Evaluation solution eliminates learning effects by counterbalancing the order in which the panelists evaluate the Web sites. Learning effects, also known as order bias, occur when panelists learn to do tasks better on the second Web site and thus do not accurately portray the usability of the second Web site. In order to eliminate potential learning effects, Keynote reverses the order of the sites for half of the panelists. Thus, Keynote can attribute any differences in customer experience to differences between the sites and not to the order in which panelists evaluated the sites.

Sample Size for Within-Subjects Designs

Comparative Evaluations employ a within-subjects design such that each panelist evaluates both sites, which eliminates error variance due to individual differences. Although reduced error variance means that even a small sample size is likely to detect statistically significant differences between the sites, Keynote recommends a sample size of 200. The sample size is necessary for meaningful descriptive statistics of market research data to inform strategic business concerns, and to do between-subjects analyses on the first-site data when there are order effects.

Previous Exposure Effects

Previous exposure effects are similar to learning effects. Panelists who have had previous experience with a Web site in their personal history may evaluate a site differently from someone who is less familiar with it. For example, it would not be surprising that success rates show a leading site as more usable than a less established site, even if the lesser-known site were more user-friendly. Web users may be more experienced with the leading site's procedures and find them easier to accomplish. Panelists can be asked to report their familiarity with the site to determine whether previous exposure may be a factor in their site preferences. Controlling for previous exposure in constructing samples (e.g. selecting panelists who have equal levels of experience with all the sites evaluated) must be weighed against the need to realistically measure brand awareness among competing sites.

Between-Subjects Designs

Clients can use Keynote evaluations to compare different versions of the same Web site or compare different Web sites, creating a true experiment with random assignment to alternative design solutions. For example, a client may want to compare two prototype registration processes, or compare two different sites' registration processes. Keynote can send a sample of panelists to different versions of the same site with the different registration processes. This method leads to definitive answers as to which design is most effective.

Cross-Sectional and Longitudinal Designs

Web sites can be compared before and after design changes to ensure that design changes produced the intended improvements in customer experience. In a cross-sectional approach (different individuals but similar sample parameters at two different points in time), or alternatively,

The Keynote Evaluation Development Team

Keynote evaluation methodology is developed and refined by consultants, who work directly with clients, as well as with a team of research experts who conduct internal evaluations.

Keynote Consultants

Keynote supplements its core offering through the work of its Professional Services division, which includes a team of in-house consultants. Keynote consultants add value to client projects by guiding evaluation strategy and design, assisting in the interpretation of results, and making recommendations based upon those findings. Consultants have training in both market research and usability methods, as well as in advanced Keynote evaluation practices. Each consultant has extensive experience developing evaluations for a variety of clients. Their collective experience informs the research team, to allow a continual fine-tuning of the methods and cataloging of best practices and benchmarks.

Keynote Research Team

The Keynote research team defines and refines the methodology, conducts primary research on best practices and compiles industry benchmarks. This team conducts internal research and stays abreast of scientific developments in measuring customer experience to continually refine the Keynote methodology. In addition, the research team analyzes aggregate results across evaluations to identify best practices and establish benchmarks for comparison with particular evaluation results.

Some of the key members of Keynote's research team include:

Dr. Anthony Bastardi is Senior Research Scientist at Keynote overseeing quantitative analytics. He is an experimental psychologist with over 15 years of experience conducting theoretical and applied research in cognitive and social psychology, specializing in statistics. His work has been published in leading academic journals and includes research on behavioral decision-making, attitude and belief formation, information pursuit and use, and strategic behavior. He has served as Research Associate in the Woodrow Wilson School of Public and International Affairs at Princeton University where he conducted research exploring motivational influences on the interpretation and evaluation of web-based information relevant to social issues. He received a M.S. in Statistics and a PhD in Experimental Psychology from Stanford University where he worked with Dr. Lee Ross and Dr. Amos Tversky.

Dr. Bonny Brown is Director of Research at Keynote. As Director of Research, Bonny is responsible for overseeing the advancement of the research methodology that forms the basis for the Keynote solution. She also leads the panel services, educational services, and the research and statistical consulting teams at Keynote. She has more than 15 years of experience in both qualitative and quantitative research in psychology and human computer interaction. Bonny's published work includes a range of topics in psychology, with an emphasis on situational influences on motivation and emotion. She is co-founder and president of the Bay Area chapter of the Usability Professionals Association (UPA), and has conducted extensive primary research on usability and marketing research methodologies. Prior to joining Keynote, Bonny worked as a Research Scientist at the American Institutes for Research (AIR) Cognitive Labs and Center for Community Research and taught and conducted research at Stanford University. She received her B.A. from University of California, Berkeley, and both her M.A. and Ph.D. from Stanford University, where she worked with Drs. Mark Lepper and Robert Zajonc.

Summary of Keynote Methodology

The Keynote methodology takes a new approach to customer experience testing by combining the critical elements of traditional usability testing, market survey research, and log analyses into a single evaluation in a fast, cost-effective manner. Large samples of users evaluating the site remotely provide more reliable, more representative data than traditional usability testing. The intent-based context allows for meaningful interpretation and aggregation of users' behavior, as recorded in log files. The approach has the unique advantage of relating the different aspects of customer experience, including brand positioning and expectations, users' behavior, and subjective experience of the site itself, into one analysis.

Keynote's unique research methodology is able to provide accurate, reliable data and insight for a window into the Web user's experience and how that experience might best be improved. Thus, marketing executives discover how they should allocate resources for maximum impact, while designers obtain insight into why particular features and functions are not working as planned and how to best modify them.

Appendix A: Keynote Checklist for Formulating Studies

Based on proven market research and usability methods,⁵ the Keynote evaluation development process consists of four phases: formulating strategy; developing protocol; data collection; and analysis and delivery of results.

Phase I: Formulate Evaluation Strategy

In this phase, areas of concern are identified, appropriate target markets are selected, and the best evaluation design is chosen. Specific steps are outlined below.

- Define the concerns. Gather questions and concerns regarding design issues (e.g., site look and feel, navigation, registration, purchasing) and marketing issues (brand awareness, positioning, feature requests).
- Decide what data is needed. What is already known? What more needs to be known for the next set of important decisions?
- Decide on the primary purpose of the evaluation. Is the study exploratory? Are there specific hypotheses to test? What comparisons would render the results most meaningful?
- Decide on evaluation design. Which design would best address the concerns (e.g., single-site, Comparative Evaluation, before-after, multiple conditions of the same site)? For evaluations with comparisons, decide whether a within-subjects or between-subjects design is more appropriate.
- Decide on the best user profile to achieve aims of the study. Develop a profile of this target market with respect to demographics (e.g., age, gender, income), and “webographics” (Web usage patterns, such as novice users versus power users).

Phase II: Develop Evaluation Protocol

In this phase, the team will develop a set of tasks for panelists that address objectives of the study and choose supporting questions. Tasks include:

- Following the aims of the study defined above, choose objectives (typically three to six) and supporting questions that will provide data to inform the relevant design and marketing decisions (e.g., evaluation success of registration process and ask users whether they found it frustrating). What specific tasks will reveal problems of interest? For example, to determine how easy it was to search for goods on an e-tailing site, the specific task given to the panelists might be “find a men’s red cashmere sweater.” The Keynote template library is a valuable source for finding typical objectives and related questions developed from previous evaluations.
- Anticipate how behavioral measures (e.g., click streams, time intervals, page views) will provide insight into key concerns. Behavioral measures—what customers actually do—versus their opinions and self-reports of behavior are generally viewed as the most reliable and predictive measures.
- Decide on attitudinal measures that address key concerns. What attitude ratings and open-ended comments would be most helpful in understanding why people are succeeding or failing at a task?
- Select questions that further define market segments, such as, “How frequently do you purchase clothes online?”
- Add conditional questions. Conditional questions are questions that are asked only if a prior condition has been met. For example, if a panelist says they are giving up on an objective, ask the open-ended question “Why are you giving up?”

⁵ For review of recommended survey development practices: Churchill, G. A. ((1999). *Marketing Research: Methodological Foundations*. New York: Dryden Press, Harcourt Brace College Publishers.

For a review of recommended usability test development process: Dumas, J. S. Redish, J. C. (2000). *A Practical Guide to Usability Testing*. Portland, OR: Intellect.

- Anticipate users' answers. If it can be anticipated that everyone will fail (floor effect) or that everyone will agree (ceiling effect), then a different specific task or question will probably yield more interesting (i.e., surprising) information. Generally, tasks of moderate difficulty reveal the most because there is a wide variability in users' ability to accomplish the task.
- There are typically three phases of a Keynote evaluation: the introduction, the objectives and associated questions, and wrap-up questions.
- Pre-testing is critical to make sure users understand and interpret questions as intended. Are any questions confusing? Can users get through the protocol in a reasonable amount of time? What are the expected results?
- Evaluation scripts go through the Keynote quality assurance review before being used to catch technical errors.

Phase III: Data Collection

During this phase, the evaluation is launched and data is collected. Steps include:

- Panelists meeting sampling criteria are sent email invitations to participate in a study.
- To prevent response biases, Keynote does not disclose any information about the Web site users will be evaluating in the evaluation invitation.
- Panelists receive the invitation. They follow the URL associated with the site and begin the evaluation using the Keynote Connector, a small downloadable companion to Microsoft Internet Explorer.
- As sampling quotas are met, the evaluation invitations stop.

Phase IV: Analysis and Delivery of Results

Phase four is when results are interpreted and recommendations are developed:

- Analyze evaluation results. Data are automatically compiled into CustomerScope, Keynote's online interface for examining data in easy-to-read charts, graphs, and other data aggregation tools. Data can also be exported to a flat file for additional statistical analyses using statistical software. Qualitative data can be searched using keywords or by common theme, and are also organized by group, such as by whether users succeeded, failed, or gave up at a task.
- Examine key measures and previously held hypotheses. Are hypotheses confirmed or disconfirmed?
- Based on data, develop theory for why users behaved and felt as they did.
- Check qualitative data for supporting and disconfirming evidence of this theory.
- Explore data for unanticipated insights.
- Develop action agenda based upon findings.