



## Using the Right Tools

To Measure and Manage Customer Satisfaction, Value, and Loyalty

### TOOLBOX

- On Bus Seats, Handrails & On-Time Performance: Customer Satisfaction Measurement – The Critical Link Between Customers and Transit
- The People On The Bus Go Up & Down: Determining Who to Interview
- Vanpools, Buses, & High-Speed Rail: Determining Your Vehicle – Qualitative and Quantitative Research
- Taking The High Road: Conducting Research – From Questionnaire Design to Sampling
- Raising The Hood: Data Analysis and Interpretation

### Customer Satisfaction Research: A Key to Keeping Customers

In recent years, increasing emphasis has been placed on customer retention programs. American business in general, and many transit agencies, have become increasingly committed to the idea of customer satisfaction and product-service quality as a means of retaining their existing customer bases. The measurement of customer satisfaction and its link to product-service attributes is the vehicle for developing a market-driven quality approach.<sup>xliii</sup> So important is customer research that winners of the prestigious Malcolm Baldrige National Quality Award<sup>\*</sup> are expected to use customer research to be able to describe how the company:

- Determines current and near-term requirements and expectations of customers, addresses future requirements and expectations of customers, and evaluates and improves its processes for determining customer requirements and expectations.

<sup>\*</sup> The Malcolm Baldrige National Quality Award was established in 1987 by the U.S. Department of Commerce. The annual award recognizes U.S. companies that have excelled in quality achievement and quality management. It recognizes and encourages commitment to quality products and services.

- Provides effective management of its interactions and relationships with its customers and uses information gained from customers to improve customer relationship management processes.
- Demonstrates commitments to customers regarding its products and services and how these commitments are evaluated and improved.
- Determines customer satisfaction, customer repurchase / reuse intentions, measures customer satisfaction relative to competitors, and evaluates and improves its overall processes and measurement scales for determining customer satisfaction.
- Summarizes trends in the company's customer satisfaction and trends in key indicators of customer dissatisfaction.
- Compares the company's customer satisfaction results with those of competitors.<sup>xliv</sup>

Many transit agencies are following this trend with the introduction of Customer Service Initiatives. For example,

The Metropolitan Council of Transit Operations (MCTO) in Minneapolis implemented a program called, *The Customer is the Boss* initiative. As part of this initiative, the agency uses a Customer Service Index (CSI) to gauge its performance in satisfying customers' needs from the customers' perspective.

Various approaches to customer retention research programs have been developed to meet differing requirements. No one methodology is perfect. However, over the past several years, general principles and guidelines have developed that serve as a starting point to developing a research program that achieves the desired goals. The purpose of this chapter is to discuss and illustrate some of the basic principles of setting up a successful customer research program. A **"successful" program is one that provides relevant information that can be used to plan for and develop transit services, and to develop marketing strategies targeted at achieving ridership growth by retaining existing customers.** This chapter's focus is on the design and implementation of a customer research program targeted at retaining existing riders. An effective customer satisfaction research program often is an integral part of this overall effort.

This chapter presents the basic tools for conducting a customer satisfaction measurement research study. While the focus is on customer satisfaction research, much of the material is applicable to any market and customer research effort.

## The Basics of Customer Satisfaction Measurement

On the surface, the task of conducting customer satisfaction research seems simple. Transit agencies conducting customer satisfaction research typically have taken a report card approach to measuring customer satisfaction, usually conducting on-board surveys or intercept interviews at major points of departure and arrival. Evaluations of performance quality are collected periodically, often while completing other research such as origin and destination studies. Results from this research provide an indication of improvement or deterioration since the last measurement.

The primary limitation of this approach is that it does not provide a clear indication of what improvements to make. For example, an agency may find itself receiving poor ratings on one attribute – for example, inside cleanliness of buses – and move rapidly to make improvements in this area. Results from the next wave of interviewing several years later show that there has been no change in overall ratings or satisfaction with the system. Faced with this situation and in light of declining budgets, transit agencies often find themselves questioning the value of this customer research. Rather than questioning the design and methods used, agencies find themselves using the research less and less, and over time missing reporting periods or delaying reports until the information is no longer timely.

What seems a simple process – asking customers to evaluate an agency's performance – actually is much more complex in practice. Without proper assistance, agencies often find:

- A lack of discrimination in customer responses,
- Highly skewed results,
- Unstable findings that vary randomly between reporting periods,
- Inability to determine the factors driving satisfaction,
- Substantial amounts of missing data,
- Meaningless or unimplementable conclusions (e.g., add more routes),
- Interpretation problems,
- Lack of organizational buy-in, and
- Nonuse of study results.

Those who consider customer satisfaction measurement as a simple report card type survey are not realizing the full value of customer satisfaction research as a management information system. To begin to understand the full value of a customer satisfaction measurement system one must begin with a definition:

Customer satisfaction measurement is a management information system that continuously captures the voice of the customer through the assessment of performance from the customer's point of view. Customer satisfaction research measures external, or performance quality.

An effective customer satisfaction measurement system must be able to determine how best to improve customer perceptions of product and service quality. Actionable customer satisfaction research provides the information that transit managers need to make changes in the processes that affect customer perceptions of service quality. The purpose of this chapter is to outline some basic tools and methods that can be used to guide and improve the customer satisfaction measurement process. Central to the effectiveness of a customer satisfaction measurement program is the accomplishment of the following steps:

- 1) Identify the market segments, customer and potential customer groups that are most likely to be affected by service quality improvements.
- 2) Determine the critical performance attributes that result in customer satisfaction.

- 3) Develop a research design and methodology that yields reliable and statistically valid data and analyses upon which to base business decisions.
- 4) Assess the performance of the agency.
- 5) Demonstrate the relative impact of the various satisfiers and dissatisfiers on overall perceptions of service quality at an agency.
- 6) Identify actions that will lead to increased satisfaction.

The following Roadmap highlights criteria for improving the process of customer satisfaction measurement. This chapter identifies basic tools to achieve these improvements.



## **ROADMAP 7 CUSTOMER SATISFACTION MEASUREMENT CRITERIA**

- ✓ **Coverage**..... of markets, customer populations, products and services, competition.
- ✓ **Frequency**..... of measurement, from ad-hoc to periodic to continuous.
- ✓ **Flexibility** ..... to adapt to unique market and customer circumstances.
- ✓ **Validity** ..... of data – data collection methods, questionnaire design scales, etc.
- ✓ **Representativeness**..... of data – sample design and nonresponse bias issues.
- ✓ **Thoroughness**..... of data – the degree to which all aspects of performance quality are measured, the use of both qualitative and quantitative data and information, and the nature and extent of comparative information (benchmarks).
- ✓ **Specificity**..... of information – actionability.
- ✓ **Precision**..... of data – confidence levels and intervals for random samples.
- ✓ **Nature and extent** ..... of data analyses – relevance, validity, usefulness, frequency.
- ✓ **Involvement**..... of employees and management in research design.
- ✓ **Continuous improvement**.... of research process.

## **Who to Interview**

At first glance, most organizations conduct their research on customers, who are defined as their existing users. This narrow focus, however, frequently results in a limited view of the environment in which a transit agency operates and may severely limit an agency's ability to take a proactive role in identifying problems and opportunities.

The requirements for the Malcolm Baldrige National Quality Award emphasize the importance of identifying customers:

Identify the market segments, customer and potential customer groups, including customers of competitors, and their requirements and expectations through surveys, interviews, and other contacts.

Defining customers can be difficult and is one of the major challenges in customer satisfaction research. Just who are customers of public transportation? Current riders are certainly customers. But, what about former riders? Former riders are past customers and represent potential customers. What about nonriding members of a household in which a rider lives? These individuals may be influential in determining current and future mode choice. As such, they also represent valuable sources of information about transit performance. Can staunch nonriders – those individuals who will never use public transportation – be considered customers? As indirect beneficiaries of a good public transportation system – improved quality of life, less congestion – and as taxpayers, they may also represent an important "customer" base.

Even when customers are defined simply as existing riders, the decision regarding who should or should not be surveyed is not straightforward. Transit agencies use many definitions of their existing riders. For example,

Metro Transit in Seattle defines a rider as anyone who has taken five or more one-way rides on a Metro bus outside their downtown Free Ride Area in the last thirty days.

MCTO in Minneapolis defines a rider as anyone who rode any MCTO bus in the last year. Milwaukee Transit has defined a rider as anyone who has ridden the bus at least once in the past thirty days.

Differences in defining riders means that comparisons of results for surveys conducted by different transit agencies should be done cautiously.

Interviews among different populations provide different types of insights. We directly relate the decision, then, whom to interview to the decisions facing the agency and the overall purpose of the research (see chapter 4). Interviews with current riders give agencies the central perspective of the entire customer satisfaction measurement program. However, interviews with other groups can be equally important. Interviews with former riders may reveal perceived quality barriers to use, pinpoint fatal flaws, and identify the benefits that customers perceive in switching modes. For example,

Houston Metro found that most former riders are not making the same trip now that they had used Metro for in the past. A failure by the agency to keep up with these former riders at a point when they were making a mode choice decision resulted in significant rider attrition.

Finally, research with nonriders can help establish an industry norm or benchmark. How the market at large perceives public transit or rail compared with other mode choices is perhaps the key element of strategic positioning.



## ROADMAP 8 WHO TO INTERVIEW

✓ <b>Current Riders</b>	<ul style="list-style-type: none"><li>• Principle population of interest</li><li>• Final judge of quality</li></ul>
✓ <b>Former Riders</b>	<ul style="list-style-type: none"><li>• Sources of dissatisfaction</li><li>• Sources of competitive intelligence</li></ul>
✓ <b>Nonriders</b>	<ul style="list-style-type: none"><li>• Principle source of competitive intelligence</li><li>• Necessary for benchmarking</li></ul>
✓ <b>Stakeholders</b>	<ul style="list-style-type: none"><li>• Sources of public support</li><li>• Necessary for benchmarking</li></ul>

Typically when deciding who to interview, focusing on a relatively homogeneous population of customers in any one study is best. For example,

MCTO (Minneapolis) conducts a quarterly study focusing solely on current riders. Also for the past two years, they have undertaken a second research effort, focusing solely on nonriders.

By limiting the population of interest to a particular customer type, the results will be more meaningful and actionable than if the research includes multiple customer groups. When the research includes multiple customer types, attributing study results to an identifiable segment of customers that can be singled out for appropriate action can be more difficult. Finally, an additional advantage is that more interviews with only one type of customer group makes it possible to generalize sample results to this population with greater confidence.

However, it should be noted that there are some advantages to interviewing different groups rather than a homogenous population. By determining key points of attitudinal and behavioral differences between groups, one can gain a better understanding of the possible reasons behind these differences.

## Determining Critical Performance Attributes

A critical component of customer satisfaction research is concerned with determining the extent to which existing products and services meet customers' needs, wants, and expectations. Meeting customer expectations results in satisfaction. Exceeding expectations may create strong customer loyalty. Conversely, not meeting expectations results in dissatisfaction and potentially rider attrition.

We formalize expectations as a set of attributes that capture the important issues by which customers evaluate a product, service, or company. These sets of attributes will differ by industry and by company. However, research in academia and industry has identified some broad dimensions on which performance is almost universally judged.

For example, researchers Parasuraman, Zeithaml, and Berry have concluded that service quality can be described adequately using five dimensions:

- **Reliability:** The ability to perform the promised service dependably and accurately.
- **Tangibles:** Appearance of physical facilities, equipment, personnel, and communication materials.
- **Responsiveness:** Willingness to help customers and provide prompt service.
- **Assurance:** Knowledge and courtesy of employees and their ability to convey trust and confidence.
- **Empathy:** Caring, individualized attention the organization provides its customers.<sup>xlv, xlvi</sup>

A variety of industries in both the private and public sector has used these dimensions effectively. They have proven particularly applicable in the public electric and gas utilities environment and they prove equally applicable to transit as Exhibit 12 illustrates.

Within these dimensions, agencies can generate a large list of individual attributes. These dimensions are not independent of each other. For instance, facets of operator attributes and information may overlap somewhat. Moreover, the list of performance attributes has been left purposely vague and user friendly in terminology to discourage the use of this list as the appropriate list for any single transit agency.



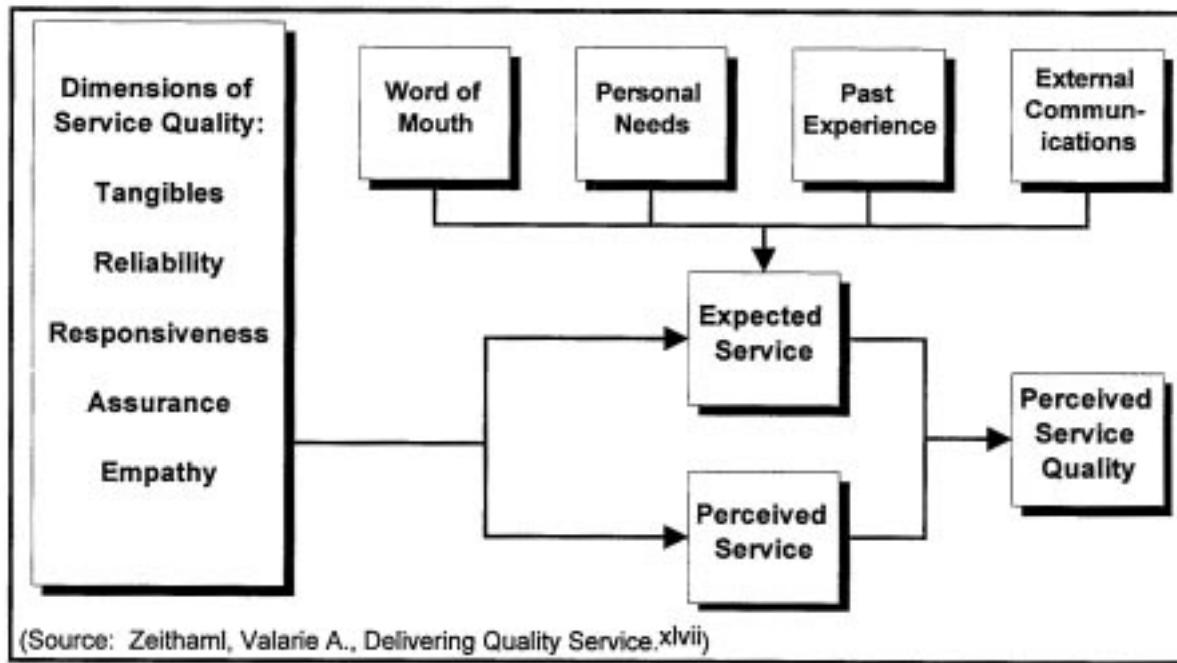
## EXHIBIT 12 DIMENSIONS OF TRANSIT SERVICE QUALITY

Dimension	Performance Attributes
✓ Reliability	<ul style="list-style-type: none"><li>• On-Time Performance</li><li>• Mechanical Dependability</li><li>• Vehicle Availability</li><li>• Operator / Driver Availability</li><li>• Other Disruptions</li></ul>
✓ Tangibles	<ul style="list-style-type: none"><li>• Operator Attributes (Physical Appearance, Professionalism)</li><li>• Comfort (On the Bus, Waiting For / Boarding the Bus)</li><li>• Plant and Equipment (Cleanliness and Appearance of Buses, Cleanliness and Appearance of Boarding / Alighting Facilities)</li><li>• Cost of Service<ul style="list-style-type: none"><li>– Prices and Fares</li><li>– Value - Price Relationship</li></ul></li></ul>
✓ Responsiveness	<ul style="list-style-type: none"><li>• Availability of Service (Convenience, Speed of Travel, Accessibility)</li><li>• Information (Communication Materials, Complaint Handling, Communications)</li></ul>
✓ Assurance	<ul style="list-style-type: none"><li>• Safety (Operational Safety, Personal Security)</li><li>• Operator Attributes (Competence and Courtesy)</li><li>• Customer Service Personnel (Competence and Courtesy)</li></ul>
✓ Empathy	<ul style="list-style-type: none"><li>• Understanding of the Customer</li><li>• Community Involvement / Social Responsibility</li></ul>

The linking of these dimensions to service quality is illustrated in the figure below:



**FIGURE 16**  
**CUSTOMER ASSESSMENT OF SERVICE QUALITY**



In determining which critical performance attributes to measure, two major guidelines should be used.

- 1) The performance attributes must be important to the customer. An attribute involving "state-of-the-art maintenance and bus storage facilities" may represent a major investment by a transit agency but be unimportant to the customer. Talking to customers is the only way to guarantee the selection of attributes that are important to the customer.
- 2) The performance attributes must be under the control and influence of the agency. Customer satisfaction research creates expectations – both internally in the organization and externally with customers – that change and/or improvements are forthcoming. Scarce resources and effort should not be spent on areas that cannot or will not be changed.

When determining the critical performance attributes that should be included as part of a customer satisfaction measurement tool, it is important to look at the issue from two perspectives – the internal or agency perspective and the external or customer perspective. A combination of qualitative and quantitative research techniques is used to identify critical performance attributes. Techniques used include in-depth interviews, Delphi techniques, focus groups, mail surveys, and telephone interviews.

## Determining Critical Performance Attributes – The Internal Perspective

Company knowledge should be the first source of information about critical performance attributes. Employees know their work and their customers. Often, they are also customers. Interviews with agency management, drivers, customer service representatives, maintenance personnel, etc. are important both to obtain information and to encourage involvement in and acceptance of the research process. The benefits of this internal exploratory research are many and will help the agency to:

- Finalize the study objectives, design, and survey questionnaire,
- Communicate more effectively with end users of the research in reports and presentations,
- Make meaningful and actionable recommendations for quality improvement,
- Make recommendations that are consistent with the agency's position, direction, strategy, and culture, and
- Achieve organizational buy-in to the program.

This exploratory research is best achieved through one-on-one interviews or in small groups. If small groups are used, care should be taken to structure groups to encourage an open discussion among peers. For example, avoid mixing supervisory and line personnel or management and union employees. The focus of these discussions is to understand the operations, administrative, and service processes that potentially influence customer satisfaction. Specifically, the agency should answer the following key questions.

- What are the agency's major customer contact areas? The objective is to identify all points at which a customer interfaces with the transit agency that are likely to represent possible "satisfiers" and "dissatisfiers" and to define the general nature of each encounter. These contact points are often referred to as "moments of truth."
- What aspects of this contact involve direct customer interaction?
- Who participates in this process on the agency side? On the customer side?
- How effectively and efficiently is this process performed?
- What are the structural and strategic impediments to better performance?

## Determining Critical Performance Attributes – The External Perspective

While company knowledge is the first source of knowledge about critical performance attributes, the next step should be to understand perceptions of organizational performance from the customers' perspective. To accomplish this goal, there is no substitute for talking directly with the customers. A variety of methodologies may be used at this point, including focus groups and/or telephone surveys.

The primary purpose of this research is to identify the "critical incidents" during the interaction between the customer and the transit agency that are the key determinants of the customer's perception of service quality and performance. "Critical incidents" are those aspects of agency performance that customers come in contact with directly. These incidents often define staff performance – helpfulness of drivers or courtesy of telephone personnel – and product quality – travel time and safe bus operation. A critical incident is a specific example of the experience of using the service or product that describes

either positive or negative performance. For example, a positive experience is a characteristic of the service or product that a customer would like to see every time he or she uses that service or product – the customer receives correct schedule and route information the first time he or she calls the agency. On the other hand, a negative experience is a characteristic of the service or product that would make the customer question the quality of the product – the bus arrives at the designated stop ahead of the posted time.

A good "critical incident" has two characteristics.

- First, it is **specific**. That is, it describes a single characteristic of the service or product – for example, personal safety while waiting for the bus. The critical incident is not specific if it describes several aspects of performance – for example, personal safety and security while waiting for and riding the bus.
- Second, a critical incident describes the service provider in **behavioral terms** or describes the service or product with **specific adjectives**. A critical incident such as "the driver was not able to help me" does not specify what the driver did and why the driver was unable to help. Instead the critical incident should state: (1) the driver was knowledgeable about routes and schedules and (2) the driver was courteous when answering my questions. The first critical incident describes a behavior of the driver and the second incident using a specific adjective to describe the service.

Two approaches are generally used for obtaining critical incidents: group interviews, and one-on-one or individual interviews. In either group or individual interviews, the method of generating critical incidents is the same. Several factors determine the success of this interviewing process.

- **Obtain ideas from people who have actually ridden the bus or train or who have had direct experience with the service for which quality is being measured** – ridesharing, telephone services, etc. Ideally, these people must be actual customers who have interacted with the service provider as they will be asked to provide specific examples of service or product quality.
- **Complete at least ten to twenty interviews.** If using group interviews, this typically translates into two focus groups. More interviews minimize the impact of insufficient information from any single respondent. If one conducts group interviews, critical incidents that are stated by one person might stimulate incidents from other group members. Because of this process, information obtained from the interviews is more likely to cover the spectrum of customer requirements completely.
- **Create a complete listing of critical incidents.** To accomplish this, the interviewer should ask each respondent to describe five to ten positive instances and five to ten negative instances of that product or service they received in the past. We can then translate these positive and negative instances into critical incidents that define good and poor service or product quality. Respondents should be encouraged to avoid using general terms.
- **Generate a list of approximately 200 critical incidents.** Many incidents will be similar to each other – "I was able to identify what time the bus arrived at the stop based on the schedule posted there" and "It was easy to figure out what time the bus would arrive at the stop where I was waiting."
- **Group critical incidents to form "customer requirements."** The key to categorizing these incidents is to focus on a specific adjective or verb they share. When one has reviewed and grouped all incidents, write a phrase that reflects the content of its incidents. This phrase

becomes the *customer requirement*. A good customer requirement contains a specific descriptive term for the service or product or a verb that describes an actual event involving the service or product.

- **Label these customer requirements with phrases or a single word describing the content of the critical incidents.** The following figure illustrates how this grouping and labeling process might occur.



### EXHIBIT 13 DETERMINING CRITICAL PERFORMANCE ATTRIBUTES

Critical Incident	Customer Requirements
<ul style="list-style-type: none"> <li>• I waited for the bus and I didn't know if I had just missed it or if it was still coming.</li> <li>• It makes me late to work.</li> <li>• Time is very important to me and I plan my schedule to meet the bus at a certain time.</li> </ul>	<b>On-time performance</b>
<ul style="list-style-type: none"> <li>• The driver didn't wait until I sat down and I nearly fell into someone else's lap.</li> <li>• The driver scares me when he drives.</li> <li>• The driver pulls up to and away from stops very quickly.</li> <li>• The driver is very considerate and makes sure that older people on the bus have a place to sit down.</li> </ul>	<b>Safe bus operation</b>
<ul style="list-style-type: none"> <li>• I can't figure out what time the bus comes to my particular stop.</li> <li>• I don't know whether I can get to where I want to go from this bus stop.</li> <li>• I can easily tell if I can connect with other buses from the bus that stops at this stop.</li> <li>• I know how much to pay.</li> <li>• I know what time buses stop running in my neighborhood.</li> </ul>	<b>Usefulness of information at bus stops</b>

The allocation process – from critical incidents to customer requirements – is central to the development of a customer satisfaction measurement tool. Having two people or two groups of people involved in this process is often effective. While having each person or group follow the same process is possible, having each person or group work in different directions can be better. Here, the first person or group follows the process described above beginning with the critical incidents, grouping them into customer requirements, and labeling them. The other group then begins with the labeled customer requirements and allocates the critical incidents to each dimension. For those requirements where there are obvious differences between groups, both groups should discuss their process and come to a consensus as to the appropriate incidents that compose a particular category.

Once the preliminary set of customer requirements is defined, one should test them with a larger group of customers. Here, the focus is on testing the reliability and validity of the measures and to test the labeling of each customer requirement as to how effectively it expresses the intent of the dimension. We often need quantitative research at this point. Either telephone or a written survey distributed on the bus and returned by mail represents good methods. Several statistical techniques can be used to select the final set of attributes. Standard reliability tests (statistical methods that evaluate the internal consistency of items in an attitude scale\*) are often used. In addition, factor analysis (a set of techniques for the study of interrelationships among variables) or regression analysis (a statistical technique that

---

\* For a discussion of statistical methods for reliability, see J. Paul Peter, "Reliability: A Review of Psychometric Basis and Recent Marketing Practices," *Journal of Marketing Research*, 16, February 1979, pp. 6 - 17.

develops an equation that identifies the extent to which one or more independent variables, [e.g., customer requirements] are related to a dependent variable [e.g., overall satisfaction]) should be considered. These analyses are very useful in guiding the final selection of performance attributes. Management may also include specific attributes, no matter the statistical significance, to formulate strategic plans.

In summary, **determining critical performance attributes is potentially the most important step in the development of a customer satisfaction measurement process**. This process also represents an excellent opportunity to obtain agency buy-in to the research effort. Roadmap 9 summarizes the process for determining critical performance attributes.



## ROADMAP 9 DETERMINING CRITICAL PERFORMANCE ATTRIBUTES

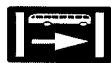
✓ Communicate the intent of the research	<ul style="list-style-type: none"><li>Publicize and discuss with employees the reasons for the customer satisfaction research.</li><li>Discuss motivation for research with departments likely to use or who would be affected by research and solicit cooperation.</li></ul>
✓ Perform internal research	<ul style="list-style-type: none"><li>Talk with the staff, and customer service personnel.</li><li>Study letters of praise and letters of complaint.</li><li>Become familiar with trends in indirect measures of customer satisfaction.</li><li>Solicit information from all levels of the organization.</li><li>Plan how the various levels of the organization will obtain and use the survey results.</li><li>Use participation to obtain a commitment to the project methodology and results.</li></ul>
✓ Perform qualitative research with customers	<ul style="list-style-type: none"><li>Define the agency's customers.</li><li>Develop discussion guides for focus groups or in-depth interviews with customers.</li><li>Conduct the research.</li><li>Develop preliminary performance attributes using the vocabulary, wording, and terminology used by customers.</li></ul>
✓ Perform quantitative research with customers	<ul style="list-style-type: none"><li>Formulate objectives using qualitative research information.</li><li>Develop the survey instrument.</li><li>Conduct the research.</li><li>Develop the final set of critical attributes from the set of preliminary attributes established during the qualitative phase.</li></ul>
✓ Review results with the organization	<ul style="list-style-type: none"><li>Review the results with management and employees.</li><li>Modify and enhance the performance attributes.</li><li>Finalize the list of performance attributes.</li></ul>

### Choosing the Right Method

We can divide customer research methodologies into two broad categories – qualitative and quantitative research. **Qualitative research** involves free-format responses to open-ended questions in which words and observations are used. Qualitative research is typically used as exploratory research. It provides in-depth information obtained from a few cases. The small number of cases cannot establish statistically reliable information for generalizations to a larger population. Two specific qualitative methodologies are in-depth interviews and focus groups.

**Quantitative research** is used to develop statistically reliable information from a sample that we can generalize to a larger population. A major objective of this type of research is to quantify the information generated from a qualitative research phase. It is used to establish the validity of preliminary conclusions drawn from exploratory research and to identify customer needs, wants, and expectations further. Quantitative research also is used to develop specific objectives, goals, and performance standards based on customer expectations. Finally, quantitative research is a key factor in developing and implementing satisfaction measures that are tracked and compared over time. Two quantitative methodologies that are often used in transit are telephone surveys and on-board surveys.

Choosing qualitative or quantitative research is not an either-or situation. The methodologies are complementary and should be combined to maximize their individual strengths. Qualitative research is often the first phase in designing customer satisfaction research, serving an important role in determining critical performance attributes. Qualitative research also can be used after a telephone or on-board survey to explain or provide additional insight into unanticipated results in the quantitative survey.



## ROADMAP 10 CHOOSING THE RIGHT METHOD – QUALITATIVE VERSUS QUANTITATIVE RESEARCH METHODS

	Qualitative	Quantitative
Type of Research	Exploratory / Conceptual	Descriptive / Factual
Types of Questions	Open-ended, interviewing process is interactive	Closed-ended, questioning is generally one-way
Number of Respondents	Few, nonrepresentative sample	Many, representative sample
Generalization of Results	Very limited	Reasonable
Data Obtained	Verbal, outcomes are subjective and open to interpretation	Numeric, outcomes precise and conclusive
Analysis	Subjective, results are an interpretation of verbal response / discussion	Statistical / objective, results are an interpretation of numerical patterns
Interviewer Qualifications	Special skills required	Less need for special skills
Cost	Higher per sample cost	Lower per sample cost
Flexibility	Easy to correct or adjust midstream	Costly to correct or adjust midstream
Time Required	Short turn-around, little lead time required	Longer turn-around, requires longer lead time as well as more time for analysis

### Qualitative Research Methods

Two methods of qualitative research are particularly applicable to conducting research on public transportation. **In-depth interviews** are face to face interviews conducted on a one-on-one basis or in very small groups. These interviews resemble conversations rather than formal, structured interviews. In-depth interviews are especially appropriate for capturing the ideas and viewpoints of key executives within an organization – the sponsoring company – or of major stakeholders in the research – firms from whom an agency contracts services or employees. The interviewers use a discussion outline designed before conducting the interviews. Questions are general and nondirective and are asked in a

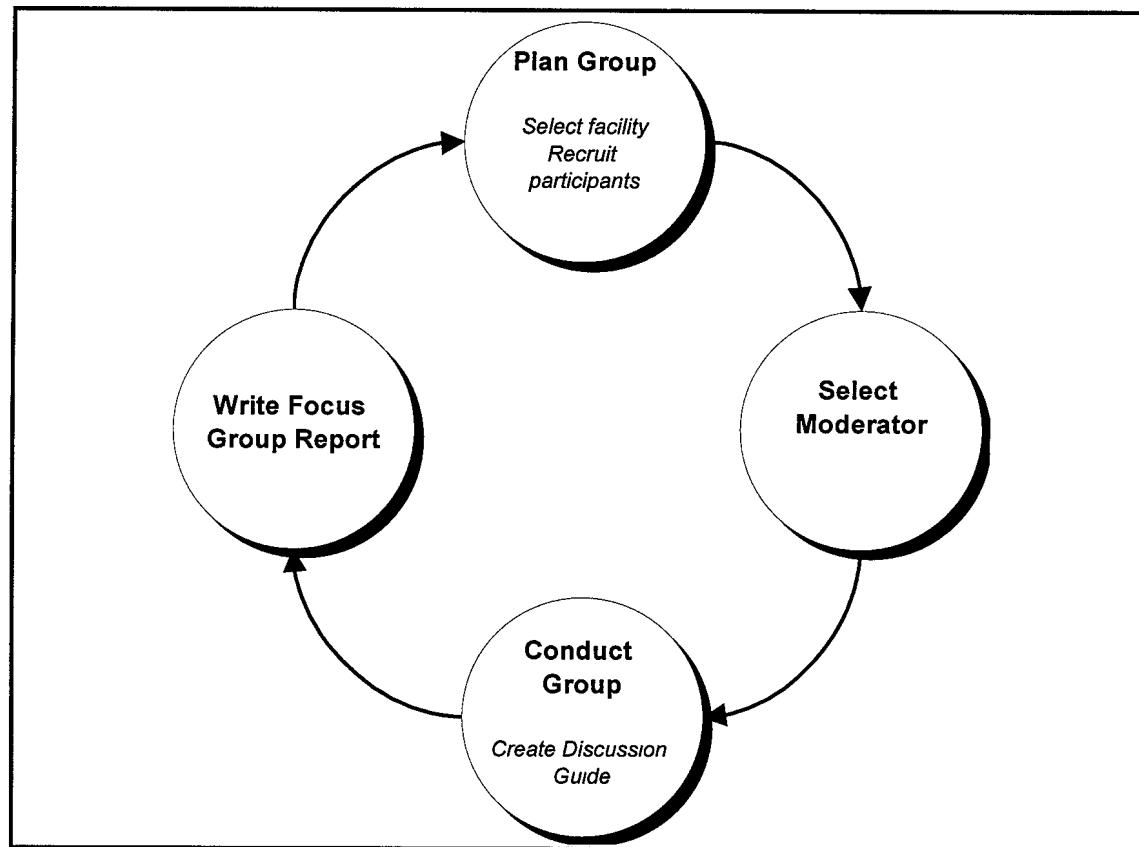
relaxed, casual manner. The interviewer must allow a free flow of ideas while still controlling the interview by providing the necessary order and structure. Time must be taken to explore and investigate important ideas. A good interviewer must be able to recognize and probe important ideas and concepts for additional information.

**Focus group discussions** are an alternative means of obtaining in-depth information. Focus groups typically are composed of eight to ten persons who share common characteristics. The groups meet for about two hours to offer opinions, viewpoints, and perceptions about a predetermined topic – for example customer requirements and expectations. Focus group sessions generate information not easily obtainable with other methods. Interactions among the group members often stimulate thinking in a manner not possible with other techniques.

Focus groups are not an easy technique to employ. A poorly conducted or analyzed focus group can yield very misleading results and waste a good deal of money. An average cost for a focus group including recruitment costs, payments to participants, facility rental, moderation, and report writing easily can be about \$3,000 to \$4,000 per group. Figure 17 illustrates the steps in planning and executing focus group sessions.



**FIGURE 17  
STEPS IN PLANNING AND EXECUTING FOCUS GROUP SESSIONS**



Four components are important to the success of a focus group:

- 1) Planning the agenda,
- 2) Effective moderation,
- 3) Recruiting correct participants, and
- 4) Analysis and interpretation of the results.

The **agenda** of a focus group typically has four distinct stages:

- 1) Introduction,
- 2) Rapport / reconnaissance,
- 3) In-depth investigation, and
- 4) Closure.

The **introduction** stage typically lasts about ten minutes and usually includes a statement by the moderator of the "generic" purpose of the session – "We're here today to talk about customer needs and expectations for transit service." The moderator also gives the ground rules for participation and makes ethical disclosures about microphones, audio and/or video taping, and observers behind the one-way mirror. This stage of the focus group ends with self-introductions by respondents. Introductions typically contain basic information about key characteristics – mode choice, transit ridership, occupation, etc. – that will help in later analysis.

The second segment – typically lasting about twenty minutes – gained the title "**rapport and reconnaissance**" because the moderator asks easy questions that build trust and uses eye contact and other bonding exercises to provide the foundation for subsequent questions. Time is spent gaining an understanding of what group dynamics are in place and what dynamics must be encouraged or discouraged to create an atmosphere that is conducive to gathering the perceptions, opinions, beliefs, and attitudes of all respondents.

Most of the group time – sixty to seventy-five minutes – is spent on **in-depth investigation**. The moderator asks specific questions and uses constructive probing techniques to find out more about respondents' perceptions, opinions, beliefs, and attitudes and to obtain reactions to specific stimuli – product / service concept statements, communication-boards, etc. During the investigation phase, there are many different ways to elicit more in-depth and insightful comments from respondents. Projective techniques are often used when respondents will not or cannot respond meaningfully to direct questions about the reasons for certain behaviors or attitudes or what the act of buying, owning, or using a product or service means to them. Participants in a group may be unaware of their own feelings and opinions, unwilling to make admissions that reflect badly on their self-image, or are too polite to be critical to a moderator. A well-trained and skilled moderator will be familiar with these methods and will know when to use them.

The final stage of a focus group is **closure**. Ideally, this time is used to conduct a summary and linking process on all that was learned in the group discussion and to obtain any last minute thoughts from participants. It normally takes about ten minutes.

The questions and possibly exercises used to guide the group are usually presented in a discussion outline. A sample guide is illustrated in Exhibit 14. This outline is strictly for general guidance and serves as a starting point for the sessions. It serves as a checklist to make certain the moderator covers all-important

topics. The set of topics covered or emphasized may change from group to group as the clients and moderator decide that a question is not generating useful, nonrepetitive information. Alternatively, a new, interesting idea may emerge, and reactions sought from subsequent groups.

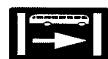


## **EXHIBIT 14**

### **SAMPLE FOCUS GROUP DISCUSSION GUIDE TO DETERMINE CRITICAL PERFORMANCE ATTRIBUTES**

<b>Introduction</b>
<ul style="list-style-type: none"><li>• Purpose of Focus Groups</li><li>• Guidelines for Discussion (e.g., no right or wrong answers, need to hear from everyone, speak one at a time, limit sidebar conversations, etc.)</li><li>• Facility (one-way mirror, people watching, audio and/or video taping, disclose for ethical reasons)</li><li>• Personal introductions (name, occupation, travel patterns, mode choice)</li></ul>
<b>Critical Performance Attributes</b>
<ul style="list-style-type: none"><li>• Thinking about your most common trip, what aspects of the trip are most important to you?</li><li>• What are some words that describe an ideal transit trip?</li><li>• What are the important factors that you consider when evaluating whether transit is a viable option for a particular trip?</li><li>• If you were talking to the head of [Agency], what would you tell him or her about [Agency's] product or service?</li></ul>
<b>Quality</b>
<ul style="list-style-type: none"><li>• How do you define quality as it relates to public transportation?</li><li>• What would you expect from a transit agency that provides superior quality?</li><li>• Describe some problems you have experienced because the "quality" of transit service does not meet your needs for your typical trip?</li><li>• Describe an instance(s) where you were very satisfied with [Agency]?</li><li>• Describe an instance(s) where you were very dissatisfied with [Agency]?</li><li>• What would you tell a friend or relative about [Agency]?</li></ul>
<b>Reliability</b>
<ul style="list-style-type: none"><li>• How do you define reliability as it relates to public transportation?</li><li>• What would you expect from a transit agency that provides superior reliability?</li><li>• Describe some problems you have experienced because the "reliability" of transit service does not meet your needs for your typical trip?</li><li>• Describe an instance(s) where you were very satisfied with the reliability of [Agency]?</li><li>• Describe an instance(s) where you were very dissatisfied with the reliability of [Agency]?</li><li>• What would you tell a friend or relative about [Agency's] reliability?</li></ul>
<b>Convenience</b>
<ul style="list-style-type: none"><li>• How do you define convenience as it relates to public transportation?</li><li>• What would you expect from a transit agency that provides superior convenience?</li><li>• Describe some problems you have experienced because the "convenience" of transit service does not meet your needs for your typical trip?</li><li>• Describe an instance(s) where you were very satisfied with the convenience of [Agency]?</li><li>• Describe an instance(s) where you were very dissatisfied with the convenience of [Agency]?</li><li>• What would you tell a friend or relative about [Agency's] convenience?</li></ul>
<b>Conclusion</b>
<ul style="list-style-type: none"><li>• If you could make one recommendation to [Agency] that would improve service in an area that is critical to your continued use of public transportation, what would that recommendation be?</li></ul>

An **effective moderator** encourages all participants to discuss their feelings, anxieties, and frustrations as well as the depth of their convictions on issues about the topic, without being biased or pressured by the situation. A trained moderator fully supports decision-makers with the most accurate and detailed indepth view of how respondents really feel about products, services, issues, and advertising. No specific degree or background guarantees a good moderator. However, experienced and well-trained moderators often have degrees, course work, or work experience in a variety of fields – psychology, sociology, anthropology, research design, marketing, business management, survey research, or retailing – or a technical background related to client work. More important than a specific degree or work history is a good grounding in group dynamics and a respect for the opinions of others. In addition, a nonjudgmental attitude and a strong foundation in a rigorous research discipline are important for a skilled moderator. Above all, an effective moderator is a good listener – one who sees and hears the research from both the clients' and the respondents' point of view.



## ROADMAP 11 KEY QUALIFICATIONS OF FOCUS GROUP MODERATORS

✓ <b>Knowledge of market research tools and techniques</b>	<ul style="list-style-type: none"> <li>Strong knowledge base about appropriate research designs for a variety of issue areas.</li> <li>Understanding of and ability to apply appropriate research tools to achieve research objectives.</li> </ul>
✓ <b>Knowledge base about the topic area</b>	<ul style="list-style-type: none"> <li>Knowledge of the subject matter in sufficient depth to ask useful questions and follow-up probes.</li> </ul>
✓ <b>Strong knowledge / experience base in group dynamics</b>	<ul style="list-style-type: none"> <li>Understanding of how to work effectively with people in groups.</li> <li>Knowledge of and ability to use more than one approach and/or solution for problems that emerge in group settings.</li> </ul>
✓ <b>Knowledge of how to analyze and report qualitative data</b>	<ul style="list-style-type: none"> <li>Ability to analyze, synthesize, and report key qualitative findings to support needs of decision-makers.</li> <li>Understanding of the limitations of qualitative research data.</li> </ul>

When **recruiting participants** for a focus group, the goal is to provide for both similarity and contrast within a single group. As a rule, combining participants from different social classes or who have very different behaviors (e.g., riders and nonriders) in the same group is undesirable, because of differences in their perceptions, experiences, and verbal skills. Also, commonality among group members avoids interactions and conflicts among group members on side issues. Within an otherwise homogeneous group, however, providing for a spark to be occasionally struck by introducing contrasting opinions is helpful. One way to accomplish this is to include a mix of participants in each group who represent otherwise different characteristics. For example, in a group of riders, one might strive for a mix of frequent and less frequent riders. Also a mix of men and women, different ages, areas of residence, etc. also can accomplish this objective. Finally, participants should be carefully screened to meet certain criteria. Participants should have adequate experience with the issues being discussed to be able to provide an opinion. People who have already participated in many focus groups or who work in particular fields – transit, market research, and advertising – should not be included.

The number of participants in a group is dependent on many factors. Although groups of eight to twelve have become customary, smaller groups may be more productive. For example, consider a group of twelve participants. After subtracting the time it takes to warm up – usually about three minutes – and the time for the moderator's questions and probes, the average panelist in a 90-minute focus group has three minutes of actual talking time. The experience becomes more like a group survey than an exploration of experiences, feelings, and beliefs.

Exhibit 15 illustrates a focus group screener. It assumes recruiting for two focus group sessions – one composed of commuters who ride the bus to work. The other group is comprised of those who drive alone. To qualify, participants must be between the ages of 25 and 64, employed full or part-time, and commute to a specific work site three or more days a week. In each group, a mix of respondents representing age and gender is desired. Standard industry and past participation screens apply.



## EXHIBIT 15 SAMPLE FOCUS GROUP SCREENER

Hello, I'm \_\_\_\_\_ with [name of research firm], a public opinion research firm in [city]. Today / Tonight we are conducting a study on transportation issues in your area. Let me assure you that this is not a sales call. This study is being conducted for research purposes only. For this survey, I need to speak with someone in your household who is between the age of 25 and 64. Would that be you?

- 1 YES [CONTINUE]                    2 NO [ASK TO SPEAK WITH THAT PERSON]  
9 NO ONE IN-HOUSEHOLD THAT AGE [THANK AND TERMINATE]
1. Which of the following best describes your employment status at this time?  
1 Employed full-time                    2 Employed part-time  
3 OTHER [THANK AND TERMINATE]
  2. Do you commute to the same work site at least 3 days a week?  
1 YES                                    2 NO [THANK AND TERMINATE]
  3. How do you usually get to work or school?  
1 Drive alone [GROUP #2]            2 Bus [GROUP #1]
  4. Which of the following categories includes your age?  
1 25-34                                3 45-54  
2 35-44                                4 55-64
  5. RECORD GENDER  
1 MALE                                2 FEMALE
  6. Do you or does anyone in your household or immediate family work:  
in market research?                    Yes                    No                    DK  
for [Agency name] or any other  
public transportation agency?        1                    2                    3  
[IF YES TO ANY ABOVE, THANK AND TERMINATE]
  7. Have you ever participated in a discussion group for research purposes for which you were paid for your time?  
1 Yes                                    2 NO [SKIP TO INVITATION]
  8. When was the last time that you participated in one of these groups?  
1 WITHIN THE PAST 6 MONTHS [THANK AND TERMINATE]  
2 MORE THAN 6 MONTHS AGO [SKIP TO INVITATION]

As a further part of this research, we would like to invite you to participate in a discussion group for research purposes. The purpose of the group will be to discuss public transportation issues and service strategies. Let me assure you that absolutely no attempt will be made to sell you any type of product or service. This group is strictly for research purposes and we would like to hear your honest opinions. You will be with about ten people and will be involved in an informal exchange of ideas and opinions.

The discussion will take place at [time] on [date] at [name of facility]. They are located in [city / neighborhood] at [address]. Free parking is available for all participants, and light refreshments are provided. This discussion will last about 2 hours and in appreciation for your time and opinions, you will receive a [dollar amount] honorarium at the conclusion of the discussion. May we schedule you to attend?

Following the group discussion, the moderator **reviews, analyzes, and interprets the results**. The moderator not only reports specific comments and findings but also looks for consistent responses, new ideas, concerns suggested by facial expressions and body language, and other hypotheses that may or may not have received confirmation from all participants. Analysis of focus groups focuses on identifying consistent responses across several group sessions. When differences between group sessions occur, it is possible to develop hypotheses based on the differences in between-group composition – for example, frequent riders compared with infrequent riders. Because the number of participants is small, the researcher does not report frequencies and percentages. Instead, reports typically include expressions like "most participants thought" or "participants were divided on this issue." A useful report of a group session is one that captures the range of impressions and observations on each topic and interprets them in the light of possible hypotheses for further testing. The researcher must put comments from respondents into a context so that implications are more evident. Meticulous documentation and interpretation of the session lay the groundwork for the final step – taking action.

A final question that often arises in planning for focus groups that is related to the analysis and interpretation of results is **how many groups to conduct**. The number of groups that an agency should conduct on a single subject depends on:

- 1) The nature of the issue,
- 2) The number of distinct market segments,
- 3) The number of new ideas generated by each successive group, and
- 4) Time and cost.

Because the focus of analysis is looking for consistency in responses across groups and because of the self-selection process inherent in the recruiting process, more than one group will always be needed. It is possible that even with the best recruiting and moderating, a group could be seriously flawed as to composition and results would be misleading. **As a rule, therefore, any project should begin with a minimum of two groups.** As the number of groups increases, there is a diminishing return in the value of the information that is obtained. The moderator invariably learns a great deal from the first discussion. The second group produces much more, but less is new. Usually, by the third or fourth session much of what is said has been heard before, and there is little to be gained from additional focus groups. Exceptions to this rule occur if there are distinct segments to cover – such as regional differences, ridership behavior, commuting, etc.

In summary, Roadmap 12 provides a simple comparison of the two primary qualitative research methods. Most of the limitations of qualitative research stem from the susceptibility of the results to misuse, rather than their inherent shortcomings. There is a great temptation among managers to accept qualitative research, based on small samples, as sufficient for making key management decisions, because they are so compelling in their reality. There are two inherent dangers in accepting the results of qualitative research.

- The results are not necessarily representative of what would be found in the population and should never be projected to a larger population.
- Second, there is typically a great deal of ambiguity in the results. The flexibility that is the hallmark of the method also gives the moderator great latitude in directing questions. Similarly, while writing the report, a moderator with a particular point of view may interpret the thoughts and comments selectively to support that view.

In view of these pitfalls, if this is the only research being conducted, results from qualitative research should be used with extreme care. Ideally, these results should be used strictly for insights into the reality of the consumer's perspective and to suggest hypotheses for further research.

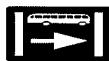


## ROADMAP 12 CHOOSING THE RIGHT METHOD – IN-DEPTH PERSONAL INTERVIEWS AND FOCUS GROUPS

	In-Depth Personal Interviews	Focus Groups
<b>Group Interactions</b>	<ul style="list-style-type: none"> <li>No group interaction. Stimulation for new ideas comes from interviewer.</li> </ul>	<ul style="list-style-type: none"> <li>Group interaction is present. May stimulate new thoughts from respondents.</li> </ul>
<b>Group / Peer Pressure</b>	<ul style="list-style-type: none"> <li>Thinking of respondents may not be challenged. Respondent may be more likely to provide unbiased or unpopular opinions.</li> <li>Role playing is minimized and no peer pressure.</li> </ul>	<ul style="list-style-type: none"> <li>Group dynamics and stimulation may clarify and challenge thinking.</li> <li>Peer pressure and role playing may occur and results may be confusing or difficult to interpret.</li> </ul>
<b>Respondent Competition</b>	<ul style="list-style-type: none"> <li>Individual is alone with interviewer and can express thoughts in noncompetitive environment. Time to obtain detailed information.</li> </ul>	<ul style="list-style-type: none"> <li>Respondents compete with one another for time to talk. There is less time to obtain in-depth details from each participant.</li> </ul>
<b>Influence</b>	<ul style="list-style-type: none"> <li>No potential for influence from other respondents.</li> </ul>	<ul style="list-style-type: none"> <li>Responses may be "contaminated" by opinions of other group members.</li> </ul>
<b>Subject Sensitivity</b>	<ul style="list-style-type: none"> <li>Respondents may be more likely to talk.</li> </ul>	<ul style="list-style-type: none"> <li>Respondents may be hesitant to talk freely or may bow to group opinions.</li> <li>Or, respondents may be willing to "admit" to an attitude or behavior if another in the group brings up the subject.</li> </ul>
<b>Interviewer Fatigue</b>	<ul style="list-style-type: none"> <li>Interviewer fatigue / boredom a problem when many interviews are needed.</li> <li>Multiple interviewers may be needed, resulting in problems with consistency with probing and analysis.</li> </ul>	<ul style="list-style-type: none"> <li>One interviewer can easily conduct several group sessions on one topic without becoming fatigued or bored.</li> <li>Only one interviewer is used, resulting in consistent probing and analysis.</li> </ul>
<b>Amount of Information</b>	<ul style="list-style-type: none"> <li>Large amount of information can be obtained, but more time is required. Costs are high.</li> </ul>	<ul style="list-style-type: none"> <li>Relatively large amount of information can be obtained in a short period of time with relatively small cost.</li> </ul>
<b>Stimuli</b>	<ul style="list-style-type: none"> <li>Large amount of stimulus material can be used.</li> </ul>	<ul style="list-style-type: none"> <li>Volume of stimulus material is somewhat limited.</li> </ul>
<b>Interviewer Schedule</b>	<ul style="list-style-type: none"> <li>Individual interviews are easier to schedule. More time is required to complete the same number of interviews.</li> </ul>	<ul style="list-style-type: none"> <li>May be more difficult to assemble eight or ten respondents at one time if they are difficult to recruit.</li> </ul>

## Quantitative Research Methods

Telephone interviews and on-board surveys are the chief methods of collecting data for customer satisfaction research in the transit industry. On-board surveys can be completed entirely while a passenger is on the bus, train, or while waiting at a station or stop. Alternatively, an agency can distribute the surveys on the bus or train, at a station or stop and request that respondents return the surveys by mail, or at central locations at stops or stations. There are many factors – time restrictions, cost, confidentiality, and the like – that determine which quantitative research method or methods should be used. Each of these two methods has its advantages and its disadvantages, all of which must be evaluated to decide which method to use. In addition, using a combination of methods may be the best decision. Roadmap 13 illustrates some key differences between methods.



### ROADMAP 13

#### CHOOSING THE RIGHT METHOD – TELEPHONE AND ON-BOARD SURVEYS

	Telephone Interviews	On-Board Surveys / Personal Interviews	On-Board Surveys / with Mailback
<b>Data Collection Costs</b>	Moderate to high	High	Low to moderate
<b>Time for Data Collection</b>	Short to moderate	Moderate to long	Moderate
<b>Control over Respondent Selection</b>	High	High	Low
<b>Response Rate</b>	Moderate to high	High	Low to moderate
<b>Ability to Access Hidden Populations</b>	Moderate	High	Low
<b>Complexity of Questions</b>	Simple to moderate	Moderate to complex	Simple
<b>Completion of Boring / Tedious Questions</b>	Moderate	High	Low
<b>Completion of Sensitive Questions</b>	Moderate to high	Low	Moderate
<b>Interviewer Bias &amp; Error</b>	Low to moderate	Moderate to high	None
<b>Ability to Ask Open-ended Questions</b>	Low to moderate	High	Moderate
<b>Perceived Respondent Anonymity</b>	Moderate to high	Low	Moderate to high

One of the primary differences in the different approaches is the direct cost of a completed interview. Exhibit 16 provides some approximate indices of the direct cost of a completed interview, to help compare data collection methods. In 1995, an index of 1.0 corresponded to a cost of \$25.00, including costs of compensation for interviewing and/or data entry personnel, training, supervision expenses, and direct costs such as telephone charges.



## EXHIBIT 16 COMPARATIVE INDICES OF DIRECT COSTS PER COMPLETED INTERVIEW

Data Collection Method	Index of Cost
• Mail Survey (costs depend on return rate, incentives, and follow-up procedures)	0.3 - 0.8
• Telephone Interviews <ul style="list-style-type: none"><li>• 7-minute interview with head of household in metropolitan area</li><li>• 15-minute interview with small segment (e.g., riders) of area population</li></ul>	0.5 - 0.8 1.5 - 2.0
• On-Board Surveys <ul style="list-style-type: none"><li>• On-board distribution with mailback or drop off at central location (costs depend on return rate, incentives, and follow-up procedure)</li><li>• 5 to 10 minute personal interview on bus or other transit location (e.g., stop, train station, or park-and-ride lot)</li></ul>	0.3 - 0.6 0.4 - 0.8

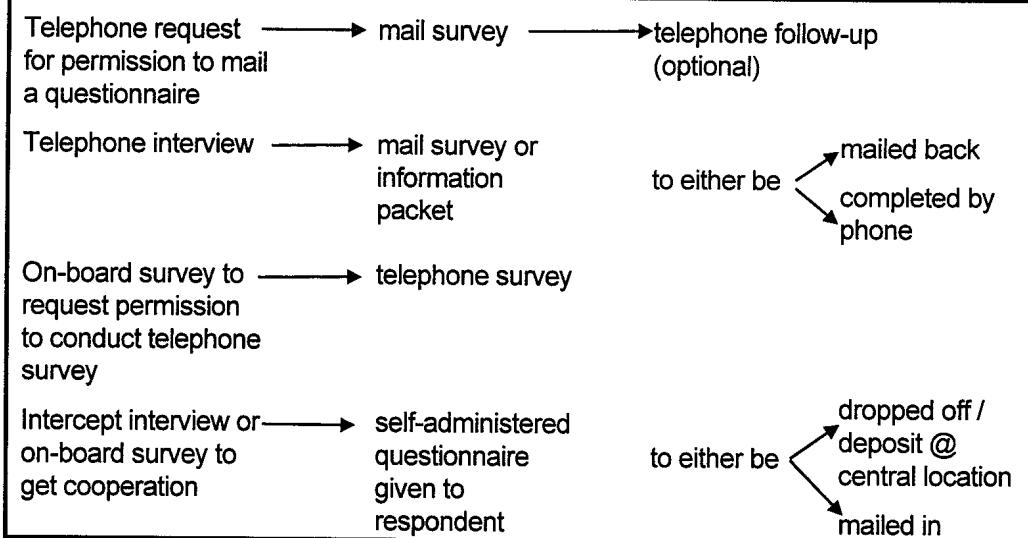
*In 1995, an index value of 1.0 corresponded to a cost of \$25.00. Costs do not include costs for research design data analysis, report writing, etc.*

## Combining Survey Methods

Since each of the basic methods of data collection has different strengths and weaknesses, it is becoming common to combine them and retain the best features of each while minimizing the limitations. Some of the feasible combinations (or sequences) are illustrated in Exhibit 17.



## EXHIBIT 17 COMBINATIONS OF SURVEY METHODS



Using telephone-interviewing methods with corresponding random sampling techniques removes much of the potential error inherent in on-board sampling methods. In some cases, it is difficult to complete the entire interview by telephone, as in where bilingual interviews are needed or where respondents need visual materials in order to respond to the questionnaire. Here, telephone interviews are used to obtain cooperation and the survey and other materials are then mailed to the respondent.

The key is the telephone presentation, which must not only gain an agreement to participate but make sure the prospective respondent is serious about cooperating. In one study in which 300 households were contacted, 88 percent agreed to complete the mail survey and 60 percent returned usable questionnaires. If the return rate is not acceptable or if analysis shows that there is a systematic bias in those who respond compared with those who originally agreed to participate a follow-up phone call should be made.

In addition, if surveys become very long or require respondents to complete some task, such as a travel diary, it may be necessary to conduct the survey in phases. For example, begin with a telephone contact and then use some other method to collect the remaining data. Seattle Metro used this method many times. Households are initially contacted and asked to complete a survey. Usually, travel characteristics, demographic data, attitudes toward issues, etc. are gathered in the first survey which can take between ten and twenty minutes to complete. Those who agree to continue are mailed a packet with information that is more complete or a description of different services being considered. In one case, these descriptions involved a conjoint or trade-off procedure useful for examining alternative service options. Respondents are recontacted by telephone to collect the data. In one such study for Metro, 812 respondents completed the initial telephone surveys. From this group 575 respondents – or 71 percent – agreed to complete the second phase. These respondents were sent a packet describing different services. A total of 437 interviews were completed. In total, those respondents who completed both phases of the research spent an average of forty minutes being interviewed. Obviously, by breaking the task up Metro secured higher cooperation rates. Moreover, respondents had the opportunity to study the service descriptions and were able to provide more knowledgeable responses to the follow-up questions. Demographic characteristics were compared between respondents who completed the first and second phase of the research. No significant differences were found between the two groups of respondents.

Transit agencies – notably those with a relatively low incidence of riders in the general population – often find themselves relying on on-board surveys to reduce the cost of conducting market research. Extreme care must be used when conducting on-board surveys to avoid problems of sampling error that can be introduced when identifying routes and trips to be included in the sample as well as the other bias that can be introduced. This bias is significant in those cases where surveys or cards are simply put out on buses or trains to be completed entirely at the riders' discretion.

Problems with sampling error can be handled using simple random sampling procedures, or to be more efficient, cluster sampling procedures. Other bias is introduced primarily in two ways. First, "interviewers" may not follow strict sampling procedures when distributing surveys, as in failing to give surveys to individuals they may feel uncomfortable approaching. On the other hand, an "interviewer" may miss the route or trip to which they were assigned and simply take another bus to make up the missed trip. Second, nonresponse bias is introduced when those handed the surveys fail to respond. Since in many instances no data is obtained from those who decline to do the survey, it is difficult to identify the extent of non-response bias as well as whether the bias is in any way systematic.

## Trends in Survey Research

Computers are being used increasingly to control the administration of questionnaires. Today, many research companies have computerized the central location interviewing process – called CATI (computer-assisted telephone interviewing). Each interviewer uses a personal computer or workstation. Questions and response categories appear on the screen one at a time. The interviewer reads the questions, enters the response, and the computer skips to the next appropriate question. For example, a question might ask whether a person who presently does not ride the bus has ridden in the past year. If the answer is "yes," that person may answer one series of questions. Those that answer "no" skip to

other appropriate questions. In addition, the computer can be used to customize the questionnaire. For example, consider the situation of an individual who works and attends school. The issue in gathering travel behavior data is whether to gather data for the work or school commute. Early in the questionnaire we might establish whether the respondent primarily commutes to work or school and then the remaining questions about commute travel are customized to apply to what they do most often – travel to work or school. No longer is there any confusion whether the responses to a particular question apply to the work or school commute. While questions such as these can be handled in a traditional pen and paper interview, the computer handles them much more efficiently.

There are several other advantages to CATI. This approach eliminates the need for separate editing and data entry steps. Consequently, the time that is required to complete the research is less. In addition, computer tabulations can be run at any point in the study. Based on an initial tabulation, some changes in research design – adding or dropping questions – can be considered. Finally, management may find the early reporting of survey results helpful in preliminary planning and strategy development.

Another use of computers is CASAQ (computer-assisted self-administration questionnaires). In this type of interview, the respondent interacts directly with the computer. The respondent is asked to sit before the computer and to answer questions as they appear on the screen. This type of interview is being widely used in malls and has some applicability in transit – for example at transit centers, rail stations, or even on-board vehicles. Computer interactive interviewing has resulted in better responses from respondents and in some cases, 30 to 40 percent cost savings. Many hardware and software developments – for example, pen computers – continue to make this form of interviewing more feasible.

Finally, the future will see a greater use of the Internet to conduct surveys. At this time, use of the Internet is limited. At its most basic level, marketing research on the Internet offers faster, cheaper ways of collecting data, and, in many cases, will provide a highly targeted list of respondents. On the other hand, marketing research on the Internet brings with it a host of unique limitations, some of which are only just beginning to be understood.

### **Summary: The Right Method**

In summary, there are many options to consider when choosing the right method to collect information. Qualitative research – in the form of in-depth interviews or focus groups – often is used to examine attitudes, feelings, and motivations. The use of qualitative research continues to grow in popularity for several reasons. First, qualitative research is usually cheaper than quantitative studies. Second, it is an excellent means to understand the in-depth motivation and feelings of customers. Third, it can improve the efficiency of quantitative research. Qualitative research, however, is not without its disadvantages. One problem is that qualitative research sometimes will not distinguish small differences in attitudes or opinions as well as large-scale quantitative studies. In addition, the respondents in qualitative studies are not necessarily representative of the population of interest to the research, severely restricting any ability to project the results of the research.

Quantitative research – in the form of surveys – continues to be the mainstay of market and customer research, notably customer satisfaction measurement. The ability to project results to a larger population as well as being able to establish and measure against a benchmark lend a richness to the data not obtainable through qualitative research.

**The ultimate selection of a method or methods requires careful consideration of the nature of the decisions that will be made based on the research, the information required, and the types of analysis that will be done.**

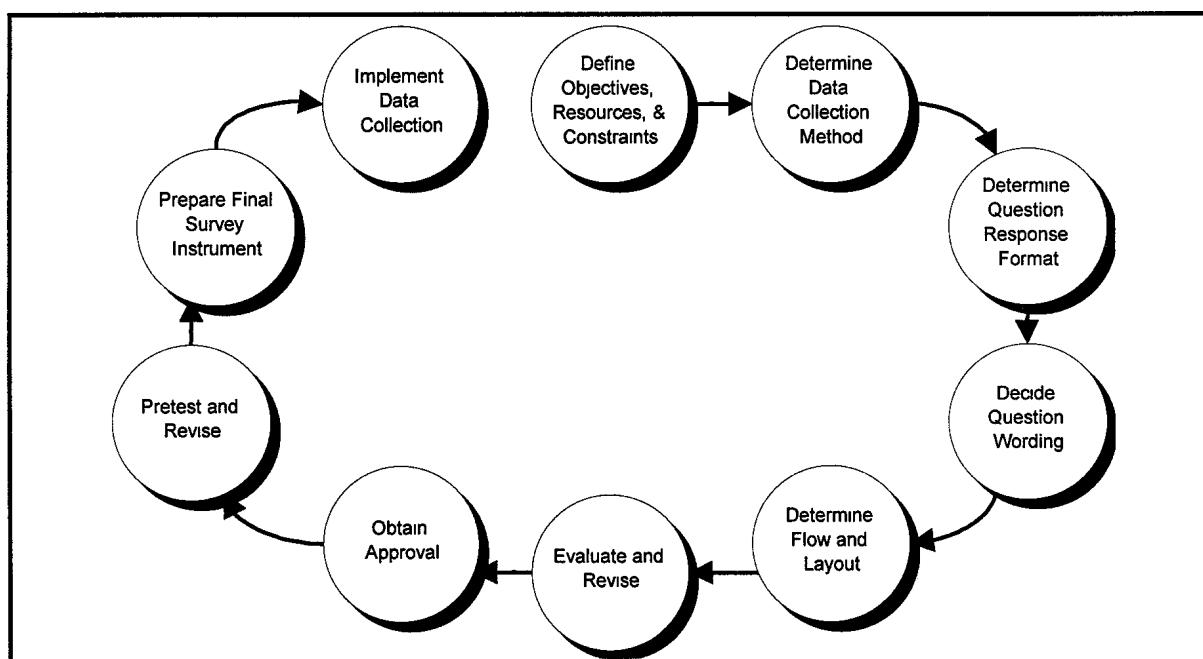
## Designing the Questionnaire

Questionnaire design is a critical phase of the customer satisfaction research process. Recently, there has been considerable discussion in the literature as to the best way to measure customer expectations and performance evaluations. Designing a questionnaire involves a logical series of steps as shown in Figure 18. The steps may vary slightly from researcher to researcher and from project to project, but all tend to follow the same general sequence. Committees and lines of authority can complicate the questionnaire design process. It is often wise to clear each step of the design process with the individual who has ultimate project authority. This is particularly true for step one – determining the decision-making information needed. All too often many man-hours are wasted on questionnaire design when a researcher develops a questionnaire to answer one type of question and the "real" decisionmaker wanted something else entirely.

It should also be noted that the design process itself - such as question wording and format – can raise additional issues or unanswered questions. Successful questionnaire development requires ongoing and open communication between the researcher and the user of the research.



**FIGURE 18  
THE QUESTIONNAIRE DEVELOPMENT PROCESS**



## Guidelines for Good Questions

The structure of questions has a direct impact on the usability of the answers. This topic is so important, there is a whole body of literature devoted to nothing else. Some excellent reference books on questionnaire design are included in the bibliography.

Survey items have two separate parts: the question and the answer. A good question not only asks for information clearly, but also elicits useful responses. Following are some general rules and guidelines to writing good questions.

- **Remember the survey's purpose.** This is key to a well-designed questionnaire. All other rules and guidelines are based on this. A test is to look at each question and ask why the question is being included. If the reason is not directly linked to the overall purpose of the research, do not ask the question.
- **Keep questions simple.** Using compound sentences and big words where simple ones will do forces respondents to remember too much information. These types of questions are likely to produce unpredictable results. One common rule of thumb is to keep the number of words in any question under twenty. If necessary, break the question down into several different parts. Using skip or branching patterns enables the interviewer to ask follow-up questions based on responses to previous questions. Use words that all respondents can understand no matter their education level, but do not sound patronizing. The most common pitfall is to use technical jargon or specialized terms. Take special care to avoid words that have different meanings for different groups. This becomes a very important issue when designing bilingual studies.
- **Keep questions focused.** Avoid words or phrases that have vague or ambiguous meanings. If a respondent can misinterpret a question, they will. A common error is not giving the respondent an adequate frame of reference, in time or space, for interpreting the question. Words such as "often," "occasionally," and "usually" lack an appropriate time reference, so respondents choose their own. Some words have many interpretations. Thus, one respondent might interpret travel time to work by bus as meaning only the time actually spent on the bus. Another might include the time walking to and waiting at the stop, the time on the bus, and the time it takes at the other end to walk to their ultimate destination. Be clear and concise. Avoid imprecise language and double negatives.
- **Avoid double-barreled questions.** These are questions in which a respondent can agree with one part of the question but not the other. Or the respondent cannot answer the question at all without accepting a particular assumption. An analyst cannot interpret answers to these questions. For example, consider this question: "Are you satisfied or dissatisfied with personal safety while riding or waiting for the bus?" If a respondent says they are dissatisfied, one does not know if they are dissatisfied with personal safety while riding the bus, while waiting for the bus, or both.
- **Avoid leading or loaded questions.** A leading question is one that clearly suggests or reveals the researcher's (or an interviewer's) opinion. For example, consider this question – "Studies have shown that congestion has become a major problem in our area. Do you agree that HOV bypass lanes are an effective solution to this problem?" While this obviously is a leading question, other examples are subtler. One way to load a question is through failure to provide a full range of alternatives. For example: "If you do not have a car available to drive to work one day, would you ride the bus or what?" On the other hand, questions can be loaded by providing a reason for an alternative. For example: "Should taxes be increased to improve public transportation and provide more service, or should we keep them about the same?" Finally, using emotionally charged words – "environment," "taxes," "tolls" – that have strong positive or negative connotations overshadow the specific content of questions.

## Getting Usable Responses

While writing good questions is important, it is equally important to pay attention to obtaining usable responses. Again, there are some general rules and guidelines to follow.

- **Make response options mutually exclusive and exhaustive.** This is the most important rule to follow when providing response options. If response options are not mutually exclusive, the respondent will have more than one legitimate place for their answer. For example: "When did you stop riding the bus? Less than one year ago? One to two years ago? Two to three years ago? Three to five years ago? Five or more years ago?" One must also ensure that the response options cover every possibility. If a complete list of options cannot be provided, include an "other" choice. If the list of choices is too long, an open-ended question might be a better option.
- **Keep open-ended questions to a minimum.** While a valuable tool, they should not be overused. Not only can they result in respondent fatigue, but they pose problems as to coding and analysis. Also, they add considerable and potentially unnecessary cost to the research. Use results from qualitative research or other studies to help generate possible response categories for open-ended questions under consideration.
- **Provide a frame of reference or other appropriate definitions for responses.** For example, a carpool is two or more passengers; a vanpool is seven or more passengers in an employer-provided vehicle. Questions that deal with time or frequency of use – using response categories such as "always," "sometimes," and "never" – are particularly troublesome. A simple definition – "By regularly, I mean at least three times a week" – will solve this problem.
- **Consider how to handle "don't know" responses.** Allowing people to say they simply do not have an opinion about a topic is useful. However, there are legitimate concerns that people will opt for that choice particularly during long surveys as respondent fatigue increases. Examine each question to decide if a "don't know" response is wise. For example, if the agency needs information only from those with an informed opinion or higher interest, offer a "don't know" option. Researchers should pay particular attention to handling "don't know" responses with scale questions. Interviewer probing should be used to learn whether a "don't know" response means the respondent has "no opinion" – the midpoint on the scale – or does not have enough information or knowledge to provide an accurate opinion – a true "don't know" response. Finally, occasionally, distinguishing between a "don't know" response and a refusal to answer the question may be important. For example, a question looking at how often a respondent is likely to use a new service may elicit a genuine "don't know" response or a respondent simply may not want to commit one way or other — a refusal to respond.



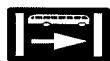
## ROADMAP 14 DESIGNING THE QUESTIONNAIRE

✓ Focus very precisely.	• Every item should zero in very directly on one, specific issue or topic.
✓ Keep each item brief.	• The longer the question, the greater the response task and the more error and bias.
✓ Strive for clarity.	• Every respondent must know exactly what's being asked.
✓ Use "core" vocabulary.	• Use the same words as the least sophisticated respondents would use in common speech.
✓ Use simple sentences.	• Two or more simple sentences are preferable to one compound sentence.
✓ Avoid specific sources of bias or error.	• Be sure items are free from the factors that create bias and error, including social desirability, prestige, threat, order, extremity, etc.
✓ Use structured questions.	• Unstructured items ordinarily provide large quantities of poor quality data.
✓ Classify answers carefully.	• Unstructured items ordinarily provide large quantities of poor quality data.
✓ Choose appropriate categories.	• Be certain they're neither too broad nor too narrow, too many nor too few.

### Using Attitude Scales

Finally, when developing survey questionnaires, particular attention should be directed to the use of attitude scales. Again, some simple guidelines are presented.

- **Provide a meaningful scale by anchoring end-points with meaningful labels.** For example: "Please rate your overall satisfaction with customer service at [agency name]. Use a scale where '1' means 'very dissatisfied' and '5' means 'very satisfied.'" Alternatively, one could give a label to each point on the scale – "very satisfied," "somewhat satisfied," "neither satisfied nor dissatisfied," "somewhat dissatisfied," or "very dissatisfied."
- **Select the length of scale – three, five, seven, or ten – that captures the greatest amount of variance in the data but is not so long that only certain parts of the scale are used.** When the measurement of extreme opinions is critical, use a scale with more points. On the other hand, generally little is gained by having a scale with more than seven points. Typically, larger scales are collapsed when it comes time to analyze the data.
- **Include a midpoint.** While there are some instances where a midpoint is not desirable, as with don't know responses it is often a good idea to give respondents with moderate or neutral opinions a way out of a forced and inaccurate response.
- **Pretest scales.** Always conduct a pretest of attitude scales paying particular attention to how well you have met the above guidelines.



## ROADMAP 15 CREATING EFFECTIVE SCALES

✓ <b>Keep it simple.</b>	<ul style="list-style-type: none"><li>Given a choice between a very short, concise scale and a more sophisticated one, use the less complex one.</li></ul>
✓ <b>Respect the respondent.</b>	<ul style="list-style-type: none"><li>Use scales that will make it as quick and easy as possible for respondents. This will reduce nonresponse bias and improve accuracy.</li></ul>
✓ <b>Choose the range.</b>	<ul style="list-style-type: none"><li>Categories and scale increments should be about the same breadth used by respondents. Normally, respondents classify things into a range from about 2 to about 7 or 8 categories, and seldom more than 10.</li><li>Do not expect respondents to be more precise than they are able.</li></ul>
✓ <b>Group only when required.</b>	<ul style="list-style-type: none"><li>Never put things into categories when they can easily be expressed in numeric terms.</li><li>Data can be grouped during processing, but if obtained in broad categories, it can't be desegregated later.</li></ul>
✓ <b>Handle neutrality carefully.</b>	<ul style="list-style-type: none"><li>If respondents genuinely have no preference, they'll resent the forced choice inherent in a scale with an even number of alternatives.</li><li>If feelings are especially strong, an odd number of scale points may result in fence-riding or piling on the midpoint, even when some preference exists.</li></ul>
✓ <b>State instructions clearly.</b>	<ul style="list-style-type: none"><li>Use language that's typical of the respondents. Explain exactly what the respondent should do and the task sequence they should follow.</li><li>List the criteria by which they should judge. Use an example or practice item if necessary.</li></ul>
✓ <b>Pilot test scales.</b>	<ul style="list-style-type: none"><li>Conduct a pretest to test instructions, format, vocabulary, etc.</li><li>Conduct analysis – using reliability and/or other statistical methods – to examine the reliability and validity of the scale items and to determine the appropriate number of scale points.</li></ul>

### A Special Case: Scales for Customer Satisfaction Measurement

The previous rules and guidelines are appropriate when developing any survey instrument. Specific to the issue of customer satisfaction research is the response scale customers would use to express their views about the performance of a product or service. The choice of a rating scale not only affects the reliability and validity of the findings, but also influences how results are used and how easy the survey is to answer and administer. Too often, agencies "borrow" or make up a scale without regard to its effectiveness as a measurement tool, resulting in biased or misleading results. What then are the characteristics of a "good" scale? A good rating scale has the following characteristics: minimal response bias, easily understood and interpreted by the respondent, discriminating power, easy to administer, and credible and useful as a management tool. Exhibit 18 illustrates some commonly used rating scales in customer satisfaction measurement.



## EXHIBIT 18 COMMONLY USED RATING SCALES<sup>xlviii</sup>

<b>Satisfaction Scale</b>			
<b>2 – Satisfaction</b>	<b>4 – Satisfaction</b>	<b>5 – Satisfaction</b>	<b>5n – Satisfaction</b>
1 = Dissatisfied	1 = Very Dissatisfied	1 = Very Dissatisfied	1 = Very Dissatisfied
2 = Satisfied	2 = Somewhat Dissatisfied	2 = Somewhat Dissatisfied	2 = Somewhat Dissatisfied
	3 = Satisfied	3 = Somewhat Satisfied	3 = Neither
	4 = Very Satisfied	4 = Satisfied	4 = Satisfied
		5 = Very Satisfied	5 = Very Satisfied
<b>Performance Scale</b>			
<b>4 – Excellence</b>	<b>5 – Excellence</b>	<b>5 – Expectations</b>	<b>4 – Requirements</b>
1 = Poor	1 = Terrible / Very Poor	1 = Much worse	1 = Missed
2 = Fair	2 = Poor	2 = Worse than expected	2 = Nearly met
3 = Good	3 = Just OK	3 = Just as expected	3 = Met
4 = Excellent	4 = Good	4 = Better than expected	4 = Exceeded
	5 = Excellent	5 = Much better	
<b>Non-Anchored Scale</b>			
Grade	Number		
A	0		
B	1		
C	.		
D	.		
E	9		
F	10		

There is no perfect scale, and considerable controversy has erupted recently over what scale to use.\* Several scales illustrated in the Exhibit 18 fail on several characteristics.

- The "2 – satisfaction" scale fails in terms of response bias and discriminating power. However, combining a strong scale for overall assessment and reserving this scale for a check-off list for identifying detailed problem areas can be an effective compromise when trying to control questionnaire length.
- The "4 – satisfaction" and "4 – excellence" scales both fail because of positive response bias and weak discriminating power.
- The "5 – satisfaction" or "5n – satisfaction" scales address the problems of positive response bias, but fare less well than excellence, expectations, and requirements scales in discriminating high-end performance.
- The "5n – satisfaction" is more difficult to explain in a telephone survey.
- The "5 – excellence" scales do well on all criteria. However, there are some problems with a tendency for the response distribution to be more skewed to the positive end of the scale when

\* See "Customer Satisfaction Measurement, Scales: A Weighty Debate," *Marketing Research* (Fall, 1994) for a discussion of different scaling options in customer satisfaction research.

compared with other scales. Analysts favor this scale when respondents are not likely to have clearly defined expectations before product / service experience.

- The "grade" and "number" scales suffer from inconsistency of interpretation. Moreover the "number" scale has too many categories and does not work well in telephone surveys.
- The "5 – expectations" and "4 – requirements" scales are strong on all criteria. Many companies prefer them as they link the notions of "expectations" and "performance" together. Both scales work well with self-administered and telephone surveys. However, care must be taken in using these scales where "exceeded" could be misinterpreted or not clearly associated with customer needs. Moreover, in telephone surveys, experience has shown that interviewers must repeat this scale more often than other scales.

**In conclusion, the choice of an appropriate rating scale to measure product or service quality is an important step in developing an effective customer satisfaction measurement tool.** The researcher should pay careful attention to the pros and cons of each scale under consideration. The purpose of the research, the types of analyses that are being considered, and the usefulness of the results for management must be considered. Above all, when selecting a scale, its reliability and validity are critical. Agencies should undertake adequate pretesting of the survey, including some psychometric testing of the measures after the pretest.

## Conducting the Research

Once the customers have been identified, the critical performance attributes have been determined, and the basic design of the questionnaire has been established, several more decisions must be made before the research can be conducted. These include what sampling procedures should be used and what quality control procedures should be instituted.

### Sampling

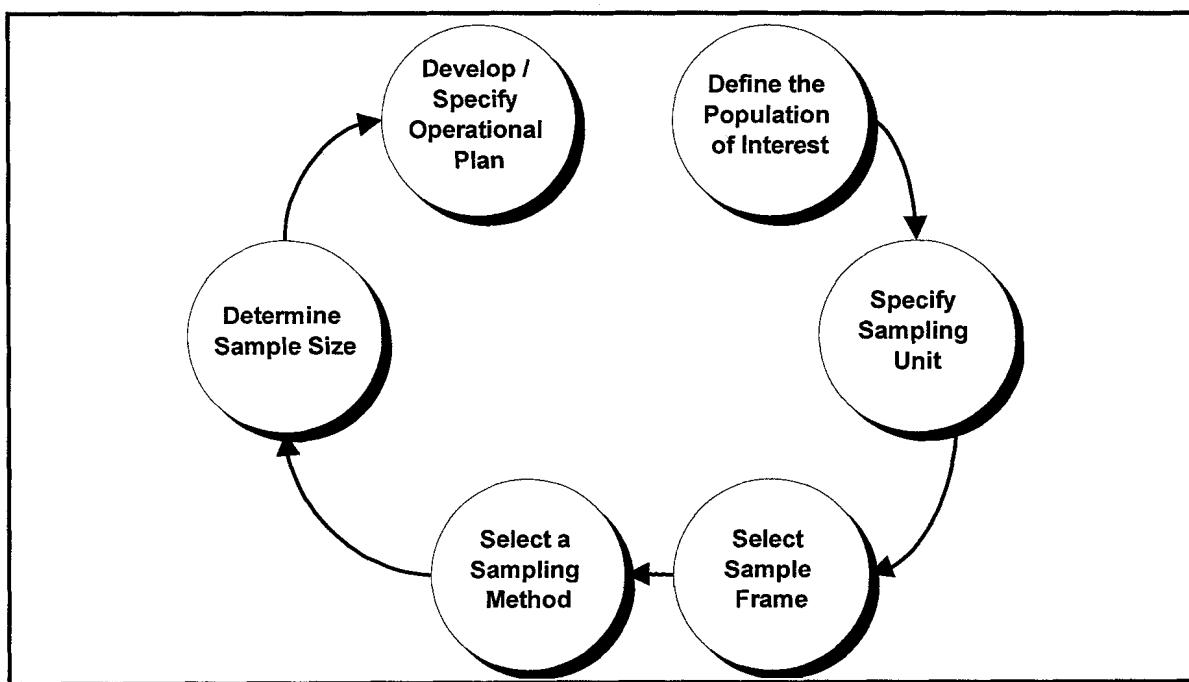
Nearly every survey research project uses some form of sampling. The main reason for sampling is economy. To survey every individual in a population is ordinarily much too expensive in terms of time, money, and personnel. Moreover, there is no need to survey every individual. By sampling a small fraction of the entire population, one can obtain estimates that represent the population as a whole with enough accuracy to base decisions on the results with a reasonable level of confidence.

While sampling is extremely practical and economical, it has to be done correctly, or it will introduce bias and error in the results. The sample must be selected properly, or it will not represent the population of interest. It has to be large enough to meet the requirements for reliability – but no so large that you are wasting resources.

Many think that the sampling decision is simply a matter of how many people to interview. Rather, many decisions need to be made. The various steps involved in the sampling process are shown in Figure 19.

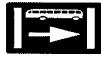


**FIGURE 19  
STEPS IN DEVELOPING A SAMPLING PLAN**



#### **Identifying the Population of Interest**

The specification of the sampling design actually begins with the identification of the **population** to be surveyed. Roadmap 16 illustrates the four main tasks involved in identifying the population of interest.



#### **ROADMAP 16 SAMPLING – IDENTIFYING THE POPULATION OF INTEREST**

- ✓ Be sure the population consists of those people who actually possess the information sought by the survey. Use your knowledge of the market to assist in this definition.
- ✓ Identify all the major factors that would otherwise qualify respondents and make their responses meaningful to the sponsor. Specify clearly who or what is excluded.
- ✓ List the criteria for inclusion and exclusion of respondents, together with the decision rules to be used. Rely on the research objectives to help define these decision rules.
- ✓ Be careful not to overdefine the population and/or that the population definition is not so restrictive that it may not be reproducible at a later point in time.

This looks like a simple task, but it seldom is. For example, transit agencies often define the population of interest as all persons living in their service area. How then does one define service area? Rarely is an agency's service area neatly defined by zip codes. Sometimes, a zip code may be only partially included. Another situation that frequently arises is where the agency serves a particular area but with extremely limited service – one route. If you survey the entire area, you are likely to reach many people who simply do not possess the information sought in the survey. The questions then are how does one exclude those individuals living in that zip code that are not part of the service area? Or, how do we target those individuals who live close enough to the route in question that they are likely to have enough information to respond?

Exhibit 19 identifies some common bases for defining the population of interest.



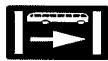
## EXHIBIT 19 SOME COMMON BASES FOR DEFINING THE POPULATION OF INTEREST

BASIS	DISCUSSION
✓ <b>Geography</b>	What geographic area is to be sampled? Usually a question of the agency's scope of operation. It could be a city, county, metropolitan area. It can be defined by zip code, census tract, or other major, easily identified boundaries.
✓ <b>Demographics</b>	Given the objectives of the research, whose opinions, reactions, etc. are relevant? Are we interested in getting information from all persons, persons between certain ages (e.g., 18 and over), or other key target groups (e.g., commuters)?
✓ <b>Use</b>	Also based on the research objectives, the population of interest frequently is defined in terms of some product or service use requirement. This is usually stated in terms of some use vs. nonuse of some quantity over a period of time (e.g., ridden the bus at least once in the past thirty days).
✓ <b>Awareness</b>	To identify persons with adequate information to respond to the questions, the agency may want to identify some screening based on awareness of agency services overall or of specific services (e.g., fixed route vs. express).

### Specifying the Sample Unit

Once the researcher has defined the population, the next step is to identify the actual sample unit to be interviewed. A **sample unit** is the smallest entity that will provide one response. Ordinarily, sample units consist of individual people. So each person in the population or one person from each household included in the population might be a sample unit. However, this is not always the case. For some surveys, the appropriate sample unit might be a household, consisting of all those living in one housing unit. This is frequently the case when travel diaries are used to gather mode choice data. Increasingly, transit has recognized the importance of employers as a target audience. Here, the sample unit may be an individual business or company.

Again, this seems a simple decision. However, consider a typical transit survey where the researcher defines the population of interest as persons living in a specified service area who have ridden the bus two or more times in the past thirty days. Here, the smallest single unit is an individual who meets these specifications. The questions then are should we interview every person in a single household that meets these specifications? Or, do we restrict our interviewing to one person per household? If the latter is chosen, what procedure do we use to select the person in the household to interview? Roadmap 17 provides some parameters for specifying the sample units.



## ROADMAP 17 SAMPLING – SPECIFYING THE SAMPLE UNITS

- ✓ Specify the sample unit so that it is the smallest single entity from which the data can be obtained.
- ✓ If the unit contains several individuals who might provide different data, the specification of the unit is too broad and should be narrowed.
- ✓ If responses from individuals would be redundant or overrepresent some entities, the specification is too narrow and should be broadened.
- ✓ If the survey data are to be compared with existing survey or secondary data, the same sample units must be used for the current survey or the data will not be comparable.

### Identifying the Sampling Frame

The third step in the process is to identify the **sampling frame**. The sampling frame is a list of the population elements or members, from which sample units are selected for sampling. In the ideal world, there is a complete and accurate list. Unfortunately, such a list does not usually exist. For example, an agency may define the population of interest for a particular study as commuters between the ages of twenty-five and fifty-four who have considered transit or another alternative mode for travel to work in the past year. There probably is no single list that provides a complete enumeration of these individuals. Sometimes, there may be several lists – persons who contacted the agency for ridesharing information, employees receiving employer-subsidized bus passes, persons who participated in a transportation fair at different work sites who requested information. While these lists combined could be used as the sampling frame, they should be carefully examined to detect what, if any, types of bias are being introduced to the study.

Often, using a larger sampling frame is simply more effective. For example, you might develop a listing of all households in the area of interest and through screening questions, determine if the person qualifies as a respondent. In a telephone interview, the sampling frame is a listing of all households with telephones. A telephone book or reverse directory may serve as a starting point for developing this sampling frame. However, the telephone book does not include those households that have unlisted or unpublished numbers.

There is substantial evidence that those with listed and unlisted numbers are significantly different in regard to several important characteristics. Research has shown that voluntarily unlisted subscribers are more likely to be renters, live in the central city, have recently moved, have larger families, have younger children, and have lower incomes than their counterparts with listed numbers. There are also significant differences between the two groups as to purchase, ownership, and usage patterns of certain products and services. If the population of interest is transit riders, this sampling frame may not include many transit-dependent riders. Unlisted numbers are more prevalent in the West, in metropolitan areas, and among nonwhites and those in the eighteen to thirty-four-year-old age group. Here, a procedure must be used to generate a list of elements of the population to be sampled. **Random digit dialing** (RDD) involves generating lists of telephone numbers at random. Many procedures exist for developing RDD samples. Also, an agency can purchase such samples from sampling companies that follow rigid procedures for developing samples and work continually to update samples, keeping track of the many changes made by local telephone companies as they add and delete prefixes and suffixes.

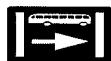
## Selecting a Sampling Method

The fourth step in the process involves the selection of a **sampling method**. The selection of a particular sampling method depends on the objectives of the study, the financial resources available, time limitations, and the nature of the problem under investigation. The major alternative sampling methods can be grouped under two categories:

- 1) Probability and
- 2) Nonprobability sampling methods.

**Probability samples** are samples selected so that every element of the population has a known, nonzero probability of selection. The simple random sample, in which every element of the population has a known, and equal probability of selection, is the best known and most widely used method. With probability sampling, the researcher must closely follow precise selection procedures that avoid arbitrary or biased selection of elements. When these procedures are strictly followed, it is possible to calculate the extent to which a sample value differs from a population value, so the researcher can project results to the population.

**Nonprobability samples** are samples where the selection of elements from the population is made in a nonrandom manner. This occurs when population elements are selected based on convenience – because they are easy or inexpensive to reach. While one can develop probability sampling methods for collecting data on-board buses, unless strict procedures are used, most on-board surveys represent nonprobability samples.



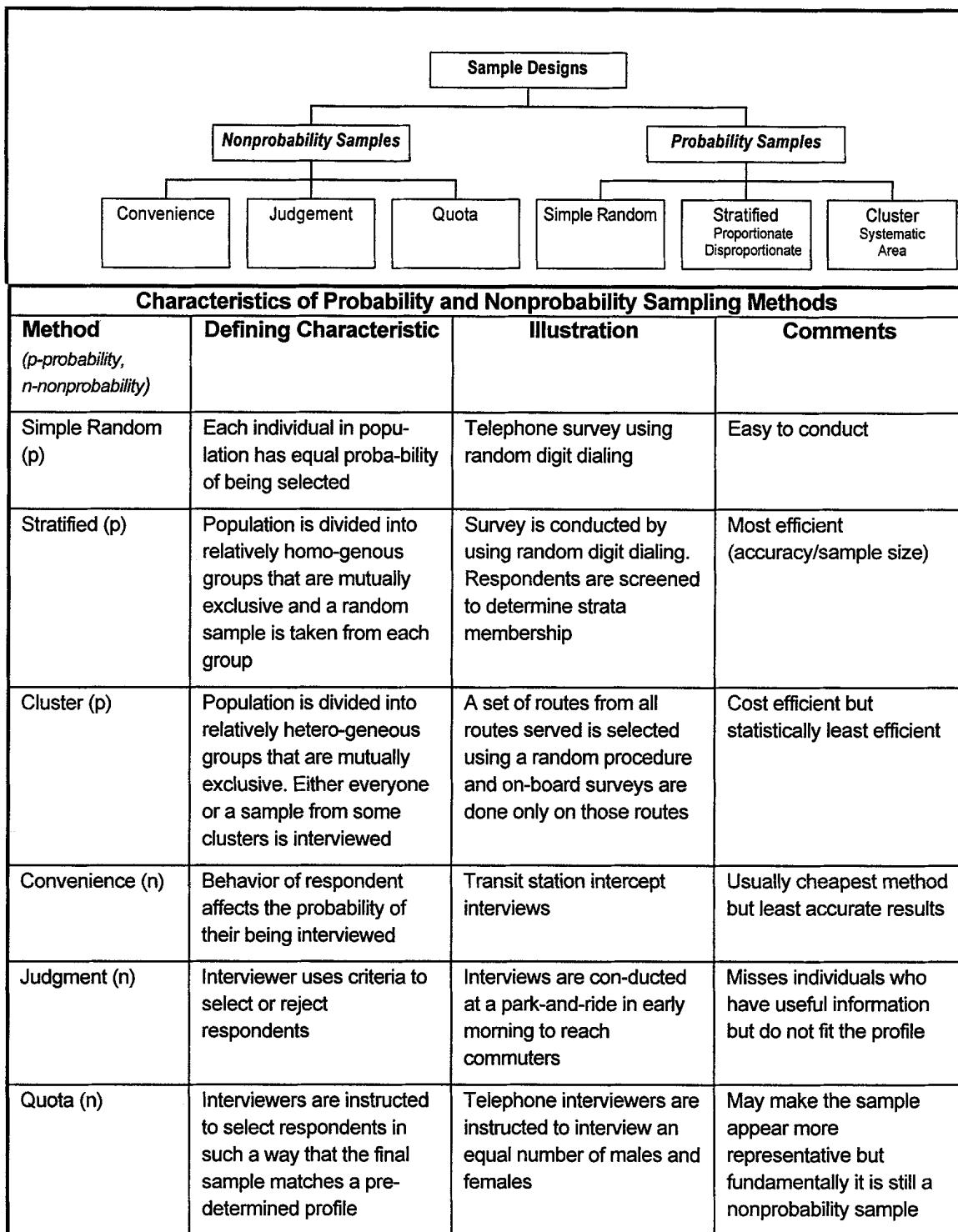
### ROADMAP 18 SAMPLING – SELECTING THE RIGHT SAMPLING METHOD

	<b>Advantages</b>	<b>Disadvantages</b>
<b>Probability Samples</b>	<ul style="list-style-type: none"><li>• Sampling error can be computed</li><li>• Survey results are projectable to the total population</li></ul>	<ul style="list-style-type: none"><li>• More expensive than nonprobability samples of the same size</li><li>• Take more time to design and execute</li><li>• Simple random samples may not be representative of the population</li></ul>
<b>Nonprobability Samples</b>	<ul style="list-style-type: none"><li>• Cost less than probability samples</li><li>• Take less time to execute</li><li>• Can be designed to achieve representative samples</li><li>• May be the only possible means of obtaining data</li></ul>	<ul style="list-style-type: none"><li>• Sampling error cannot be computed</li><li>• Results <b>cannot</b> be projected to the total population</li></ul>

Several alternative methods fall under each category. Exhibit 20 illustrates the different types of samples and their different characteristics.

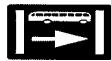


## EXHIBIT 20 CLASSIFICATION OF SAMPLING METHODS



## Determining the Sample Size

Once the sampling method has been selected, the next step is to decide the appropriate **sample size**. Many considerations come into play when determining sample size. Compromises are always being made based on the availability of resources – time, money, and personnel. Moreover, there is no one "right" way to decide what is the best or most appropriate sample size. An evaluation of the factors contained in Roadmap 19 provides some insight into whether a study needs a large or small sample. Once these factors are considered, the researcher can determine the actual sample size.



### ROADMAP 19 SAMPLING – FACTORS DETERMINING SAMPLE SIZE

<i>Factors Indicating a Large Sample</i>	<i>Factors Indicating a Small Sample</i>
✓ The decisions to be based on the survey have very serious or costly consequences.	✓ There are few, if any, major decisions or commitments to be based on the survey data.
✓ The sponsors demand a very high level of confidence in the data and estimates.	✓ The sponsors require only rough estimates concerning the parameters of the population.
✓ There is a high level of variance among the units in the population being sampled.	✓ The population to be sampled is homogeneous, with little variance among units.
✓ The sample is to be divided into relatively small subsamples during analysis and interpretation.	✓ Analysis and interpretation will be based on the entire sample, or only a few, large subsamples.
✓ Project costs and timing vary only slightly with increases in the size of the sample.	✓ A large proportion of the total project costs are for data collection or costs increase dramatically with sample size.
✓ Time and resources are readily available to cover the costs of data collection.	✓ Budget constraints and/or time requirements limit the volume of data that can be collected.

The following method for determining sample size is a conservative approach and works well for the types of research typically conducted by transit agencies. It assumes a simple random sample, a large sample approximation, and that typical sources of nonsampling error (non-response, poor administration methods, highly biased results, etc.) are trivial.

The most common approach to determining sample size is to use an equation that finds the appropriate number of cases to survey for a categorical variable – "yes" - "no" type responses.

$$\frac{Z^2(P_y)(P_n)}{C^2} = n$$

In this equation, the sample size (n) is determined by:

- Z – the number associated with the desired level of confidence;
- $P_y$  and  $P_n$  – the proportion of respondents responding "yes" or "no" to a particular question, and
- C – the number associated with the desired precision of the estimates.

To set the sample size, the researcher needs to assign values to Z,  $P_y$  and  $P_n$ , and C, and then solve for n – the sample size.

$P_y$  and  $P_n$  represent the proportion of people responding to each category in a dichotomous question (a dichotomous variable is one that has only two response categories such as "yes" and "no" or "commuter" and "noncommuter"). All variables can be converted to a dichotomous variable for this purpose. For example, if the primary purpose of the study is to identify the proportion of people in the population who ride the bus two or more times a month, a question that asks "how often did you ride the bus in the past thirty days" that has a continuous, numeric response can be considered dichotomous – a person rode less than twice or they rode two or more times.

Unless the agency has conducted research in the past, the researcher needs to estimate the probable values of  $P_y$  and  $P_n$  for the critical questions in the study. The easiest way to arrive at a number that would work for more than one question is to be as conservative as possible – 50 percent of the respondents give one answer and 50 percent give the alternative response. Thus, the computation of  $(P_y)(P_n)$  becomes simple – (.5)(.5), or .25. The equation now reads:

$$\frac{(Z^2)(.25)}{C^2} = n$$

The next thing the researcher needs to do is to decide the level of accuracy that is needed in the results and the degree of confidence one wants in the results. Most often, researchers specify a confidence interval of 95 percent. That is, they want to be able to say that nineteen out of twenty times the value of the proportion in the *population* is within a specific, numeric range from the corresponding value of the proportion computed from a sample. Similarly, an accuracy of  $\pm 5\%$  is the most common level of accuracy.

The agency should look at the nature of the decision that it is making in specifying the levels of accuracy and confidence, however. The greater the risk and the more costly a mistake will be, the greater the degree of confidence one wants in the results. In extreme cases (e.g., asking the state legislature to fund a rail system), confidence levels as high as 99 percent and accuracy levels as narrow as  $\pm 1\%$  percent might be justified. On the other hand, for low risk decisions (e.g., shifting a schedule ahead by 15 minutes), the agency may require less confidence in the results – a 90 percent confidence level with an error of  $\pm 10\%$  percent.

Decisions regarding accuracy and confidence levels directly affect the sample size. By way of example, assume that the agency would like to say that their results are accurate within a range of plus or minus 5 percent and that the agency wants a confidence level of 95 percent. The Z score is found in a table of the normal distribution. However, for a desired level of confidence of 95 percent,  $Z = 1.96$ . For a confidence level of 99 percent, the coefficient to use is 2.576; for 90 percent the coefficient is 1.67. The equation now reads:

$$\frac{(1.96^2)(.25)}{.05^2} = n$$

Solving for n yields a sample size of 384.

Larger sample sizes yield more precision. However, there are diminishing returns as sample size increases. For example, if one interviews only 100 persons, the error is plus or minus 9.8%. Increasing the sample size to 400 reduces the error to plus or minus 4.9%, but this represents a quadrupling of the sample to get a 50% reduction in the error. Increasing the sample size to 2,400 results in an error of plus or minus 2.0%. Exhibit 21 illustrates the random sample sizes required to achieve desired levels of confidence and margins of error.



### EXHIBIT 21

#### RANDOM SAMPLE SIZES REQUIRED TO ACHIEVE DESIRED LEVELS OF CONFIDENCE AND MARGINS OF ERROR

Margin of error	Level of Confidence		
	90%	95%	99%
±1 percent	6,766	9,604	16,590
±2 percent	1,692	2,401	4,148
±3 percent	752	1,068	1,844
±4 percent	423	601	1,037
±5 percent	271	381	664
±10 percent	68	97	166

The relationship between margins of error, confidence levels and sample size shown in Exhibit 21 suggests that agencies will usually opt for moderate size samples, 400-600 interviews, as this provides for reasonable error levels ( $\pm 5$  percent) with reasonable confidence (95 percent). This does not mean that one never has to use a larger sample.

In any sample size determination problem, the researcher must give serious consideration also to the number and anticipated size of various subgroups of the total sample about which one wants to make statistical inferences. For example, if the purpose of the study is to make statistical inferences only at the aggregate level – all riders in the service area – a sample size of approximately 400 is quite adequate. However, if riders living in different parts of the service area – those living in the primary service area where considerable existing service is available compared with those living in a secondary service area where there is little existing service available or new service is being added – then sample size may need to be increased to achieve reliable estimates at the subgroup level. If the agency requires the same degree of accuracy at the subgroup level, the same size would double.

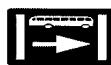
Other things equal, the larger the number of subgroups to be analyzed, the larger the total sample size required. One common rule of thumb suggests that the sample should be large enough so that there is a minimum of 100 respondents in each major subgroup. Moreover, one should strive for a minimum of thirty to fifty respondents in each of the less important subgroups.

## Quality Control

Despite the best-laid research design and sampling methods, other errors often occur in the collection of customer research information. These errors are called "nonsampling" errors. Nonsampling errors can occur because of errors in conception, logic, misinterpretation of replies, statistics, tabulation, coding, or in the reporting and interpretation of the results. They are so pervasive they caused one writer to lament:

The roster of possible troubles seems only to grow with increasing knowledge. By participating in the work of a specific field, one can, in a few years, work up considerable methodological expertise, much of which has not been and is not likely to be written down. **To attempt to discuss every way a study can go wrong would be a hopeless venture.**<sup>xlix</sup> (Emphasis added)

Not only are nonsampling errors pervasive, they are not as well-behaved as sampling error. Sampling error decreases with increases in sample size. Nonsampling errors do not necessarily decrease with increases in sample size; they may, in fact increase. Nonsampling errors also distort the reliability of sample estimates. Nonsampling errors can be reduced. But their reduction depends on improving methods and instituting strict quality control procedures. Roadmap 20 summarizes some of the most common types of nonsampling errors and some methods by which they can be reduced or controlled.



### ROADMAP 20 NONSAMPLING ERRORS AND SOME METHODS FOR HANDLING THEM

Type	Definition	Methods for Handling
<b>Noncoverage</b>	Failure to include some units or entire sections of the defined survey population in the sampling frame.	<ul style="list-style-type: none"><li>• Improve sampling frame using other sources.</li><li>• Select sample in a way to reduce incidence, such as by ignoring ineligibles on a list.</li><li>• Adjust the results by appropriately weighting the subsample results.</li></ul>
<b>Nonresponse</b>	Failure to obtain information from some elements of the population selected for the sample.	<ul style="list-style-type: none"><li>• Usually caused by not reaching those individuals not at home when attempt is made or because respondent refuses to complete study. Use strategies to address these issues specifically.</li></ul>
<b>Not-at-homes:</b>	Designated respondent is not home when the interviewer calls.	<ul style="list-style-type: none"><li>• Have interviewers make appointments.</li><li>• Call back on another day and at a different time of day.</li><li>• Attempt to contact the designated respondent using another approach (e.g., a modified callback).</li></ul>
<b>Refusals:</b>	Respondent refuses to cooperate in the survey.	<ul style="list-style-type: none"><li>• Convince the respondent of the value of the research and importance of cooperation.</li><li>• Provide advance notice that survey is coming.</li><li>• Guarantee anonymity.</li><li>• Provide an incentive for participating.</li><li>• Hide the identification of the sponsor by using an independent research organization.</li><li>• Get a "foot in the door" by having respondent comply with some small task before getting to the survey.</li><li>• Use personalized cover letters.</li><li>• Use follow-up contact at convenient time.</li><li>• Avoid interesting but not vital questions.</li><li>• Adjust the results to account for nonresponse.</li></ul>



## ROADMAP 20 NONSAMPLING ERRORS AND SOME METHODS FOR HANDLING THEM

Type	Definition	Methods for Handling
<b>Response</b>	Failure to obtain valid responses to a question or series of questions.	<ul style="list-style-type: none"><li>• Randomize question and/or response order.</li><li>• Crosscheck questions.</li><li>• Check with known data.</li><li>• Have interviewer evaluate respondent.</li></ul>
<b>Field</b>	Although the individual participates in the study, he or she refuses to answer specific questions or provides incorrect answers to them. Also, errors that arise due to interviewer error and/or bias.	<ul style="list-style-type: none"><li>• Match characteristics of interviewer and respondent as closely as possible.</li><li>• Make sure interviewer instructions are clear and written down.</li><li>• Conduct practice interview training sessions.</li><li>• Examine the interviewers' understanding of the study's purposes and procedures.</li><li>• Have interviewers complete questionnaire and examine replies to see if there is any relationship between respondents' answers and interviewer's own answers.</li><li>• Verify and/or monitor a sample of each interviewer's interviews.</li></ul>
<b>Office</b>	Errors that arise when coding, tabulating, or analyzing the data.	<ul style="list-style-type: none"><li>• Use field edit to detect the most glaring omissions and inaccuracies in the data.</li><li>• Use second edit in office to decide how to handle questionnaires containing incomplete answers, obviously wrong answers, and answers that reflect lack of interest.</li><li>• Use closed questions to simplify coding, but when open-ended questions are needed, specify the appropriate codes that will be allowed before collecting the data.</li><li>• When open-ended questions are being coded and multiple coders are used, divide the task by questions, not by data collection forms.</li><li>• Have a coder code a sample of other's work to ensure consistent set of coding criteria.</li><li>• Follow established conventions; for example, use numeric codes, not letters of the alphabet, when coding the data for computer analysis.</li><li>• Prepare code book that lists codes for each variable and categories included in each code.</li><li>• Use appropriate methods to analyze the data.</li></ul>

## Analyzing Results

Appropriate analysis of research results is critical to understanding the customers' perspectives and to developing strategies for improvement. A variety of techniques should be used to capture and analyze the important information generated from the research.

Data analysis requires the use of a set of statistical tools that reduce the amount of detail in the data, summarizing it, and making the most important facts and relationships apparent. Analyzing the results of a customer research project requires the development of an analysis plan. An analysis plan sets forth the nature and sequence of the analysis to be done. Two things dictate what statistical tools are applied or would work best:

- 1) The nature of the data – the "material" to which the tool is applied – and
- 2) The nature of the report – the "product" that is to be created.

That is, the analyst has to know what tool would work best for a particular type of data and/or report. Using the same statistical tool for every survey item is almost never appropriate.

### Simple Frequencies

When developing the sequence of the analysis, it helps to think about how the results of the data tell a story. First, it is important to understand the characteristics of the major characters in the story. The use of simple frequency tables – tables reporting the number of observations and proportion of observations in each category – provide a useful description of the variables in the data. Frequency tables also are useful to:

- 1) Determine the degree of item nonresponse,
- 2) Locate mistakes,
- 3) Identify outliers,
- 4) Determine the empirical distribution of the variable in question, and
- 5) Calculate summary statistics.

Exhibit 22 illustrates a frequency table for the overall rating of the quality of public transportation services. The question reads,

"Overall, would you say the quality of public transportation service [Agency] provides is 'excellent,' 'very good,' 'good,' 'fair,' or 'poor'?"

Both the distribution of the responses and the most commonly used summary statistics are included.



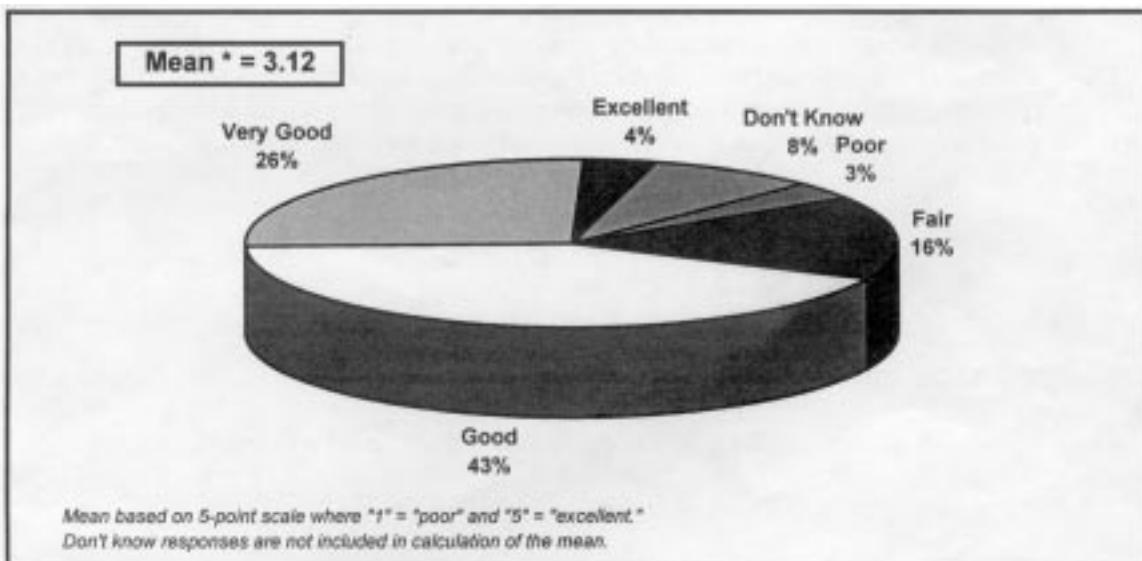
## EXHIBIT 22 SAMPLE FREQUENCY TABLE – OVERALL QUALITY OF PUBLIC TRANSPORTATION SERVICES

Value	Value Label	Frequency	Percent	Valid Percent	Cumulative Percent
1	Poor	36	3.2	3.5	3.5
2	Fair	182	16.3	17.7	21.2
3	Good	475	42.4	46.3	67.5
4	Very Good	292	26.1	28.4	95.9
5	Excellent	42	3.8	4.1	100.0
9	Don't Know	92	8.2	Missing	
Total		1,119	100.0	100.0	
Mode		3.000	Median	3.000	Mean
Std err		.027	Std dev	.867	
Valid cases		1,027	Missing cases	92	

Usability of this information is greatly enhanced through a simple graphical presentation.



## FIGURE 20 OVERALL QUALITY OF PUBLIC TRANSPORTATION SERVICES



### Cross-tabulations

The next step is to examine whether there are differences in attitudes, opinions, and behaviors of key characters or market segments. Cross-tabulation is by far the most common method to examine the associations between survey variables. It is probably used more frequently than all other methods combined. Often, analysis of survey data goes no farther than the simple reporting of frequencies and cross-tabulations. Cross-tabulation data are so common and popular in part because the method is effective, it can easily be understood and interpreted, and it can be tabulated very readily by simple spreadsheet, charting, and analysis programs. It is also very flexible and robust – that is, it will readily accept any data that can be put into a few categories. While it may lack the power and sensitivity of other measures of association between variables, cross-tabulation makes up for it by its very simplicity.

Cross-tabulation tables, or "cross-tabs," show the relationship between two categorical variables. For example, one variable could be the overall service quality rating illustrated above and the second variable could be whether the respondent is a current or former rider. Exhibit 23 shows the resulting cross-tabulation table. The table shows that there is an association between the overall rating for service quality and whether the individual is a current or former rider. This is determined by observing the distribution of responses in the various cells. For example, it is evident that the proportion of current riders who rate the agency as either Poor (2.3%), Fair (16.2%), Good (42.6%). Very Good (34.1%) and Excellent (4.9%) differ from the proportions for the Former Riders, Poor (4.3%), Fair (18.7%), Good (48.6%), Very Good (24.8%) and Excellent (3.6%). Although there is little or no difference between the proportion of current and former riders giving the agency "excellent" or "poor" ratings, current riders give the agency a higher proportion of "very good" ratings than former riders. Former riders, on the other hand, give the agency a higher proportion of "good" ratings.

The chi-square statistic is used to determine whether this pattern of results could reasonably be expected to occur because of sampling errors (i.e., by chance) or whether it represents a true difference of opinion. In this case, the chi-square value is high enough that the sampling explanation is implausible (likely to occur less than 1% of the time when there is no true difference in the opinions of the current and former riders). Consequently, the appropriate conclusion is that current riders rate the transportation services differently than former riders.

Note: One cannot properly conclude that one group is more or less favorable than the other based on the chi-square statistic alone. Furthermore, one cannot state how much more or less favorable one group is than the other. Additional analyses are necessary before making these conclusions.



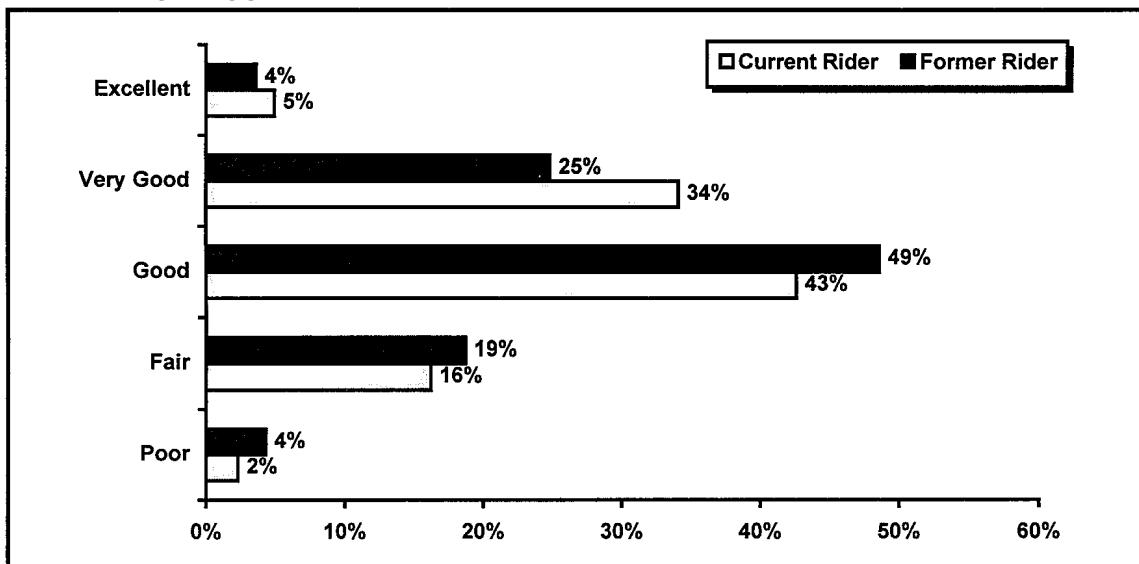
### **EXHIBIT 23 SAMPLE CROSS-TABULATION TABLE – OVERALL QUALITY OF PUBLIC TRANSPORTATION SERVICES BY RIDER STATUS**

		Count	Col. Pct.	
		Current Rider	Former Rider	Row Total
<b>Poor</b>	1	9	27	36
		2.3	4.3	3.5
<b>Fair</b>	2	63	118	181
		16.2	18.7	17.7
<b>Good</b>	3	166	307	473
		42.6	48.6	46.3
<b>Very Good</b>	4	133	157	290
		34.1	24.8	28.4
<b>Excellent</b>	5	19	23	42
		4.9	3.6	4.1
		Column	390	632
		Total	38.2	61.8
		Chi-Square = 13.60	d.f. = 4	Prob. = .00881

Again, this data can be depicted graphically and may give the user a clearer picture of the association between the two variables. Here, it is apparent that the statistically significant differences between the two groups may be attributed to the more favorable view of the current riders compared to the former riders.



**FIGURE 21**  
**OVERALL QUALITY OF PUBLIC TRANSPORTATION SERVICES**  
**BY RIDER STATUS**



While cross-tabulation tables are relatively simple to use and understand, there are several considerations when setting up a cross-tabulation table. The main requirements for using cross-tabulation tables are summarized in Roadmap 21.



**ROADMAP 21**  
**FOR USING CROSS-TABULATIONS**

✓ The objective is to determine if the distributions of one variable differ for each category of the other.	✓ If a distinction is made between an independent and dependent variable, it generally is easiest to interpret if you place the dependent variable on the rows and use the column percent.
✓ The variables to be cross-tabulated must both be derived from nominal or ordinal scales or have few enough values to be treated as categories.	✓ The chi-square statistic and its probability are used to measure the statistical significance of the relationship.
✓ One variable may be identified as an independent and the other dependent, but they need not be.	✓ The minimum expected cell frequency must be computed to be sure it is five or greater for the chi-square statistic to be valid.
✓ Either column or row percentages are most useful for interpretation, but table percentages are usually more difficult to interpret and rarely are used.	✓ If the minimum expected cell frequency is less than five, either rows, columns, or both should be combined to increase the minimum expected cell frequency.

## Quadrant Analysis

A primary purpose of customer satisfaction research is to gain an understanding of the match between customer needs, wants, and expectations and how well a company is delivering products and services that meet those needs, wants, and expectations. Quadrant analysis is one way to analyze the two dimensions of customer needs or expectations and an agency's rating simultaneously.

Quadrant analysis requires two types of data – the first identifying customer expectations or needs and the second rating an agency as to how well it meets these expectations. Asking the following two questions might gather this data.

- 1) I'm going to read you a list of items that people often consider when deciding to use public transportation. As I read each, please tell me how important this item is to you when deciding to use public transportation. Use a 5-point scale where "5" means "extremely important" and "1" means "not at all important."
- 2) Now I'm going to read the same list of items. This time, please tell me how well you feel each statement describes [agency name]. Use a 5-point scale where "5" means "describes very well" and "1" means "does not describe at all."

The items used in the questions are the same as those items identified as "critical performance attributes." There are many options for scaling. The choice depends on the type of customer being interviewed and the interpretative nature of the data wanted.

Once the researcher obtains these two types of data, the analysis is accomplished by plotting the coordinates of importance and ratings for each attribute on one graph. One axis represents the importance measure, while the other axis illustrates performance scores or ratings. The result is a two-dimensional plot that visually depicts the relationship between these two scores. This visualization enables the researcher both to analyze the data and to communicate insightful evaluations to management.

The figures on the next page show a simple, yet insightful application of a quadrant analysis. Exhibit 24 illustrates the data used in this example. While the researcher could simply present the data as shown, reading it is difficult and little interpretation is possible. For example, based on this data, an agency may elect to put its resources toward those attributes that are most important to respondents – on-time performance, nighttime safety while riding, nighttime safety while waiting, and safe bus operation. Or, they may decide to allocate resources to those areas where respondents are most dissatisfied – nighttime safety while waiting, number of stops to destination, and nighttime safety while riding.

By plotting the coordinates of importance and ratings for each attribute on one graph, the user obtains a clearer and very different picture where best to allocate scarce resources. In Figure 22, the *percent very important* and *percent very satisfied* are used. Other forms of the data – means, standardized scores, etc. – also can be used. There has been considerable literature developed as to the best types of data to use for this analysis.

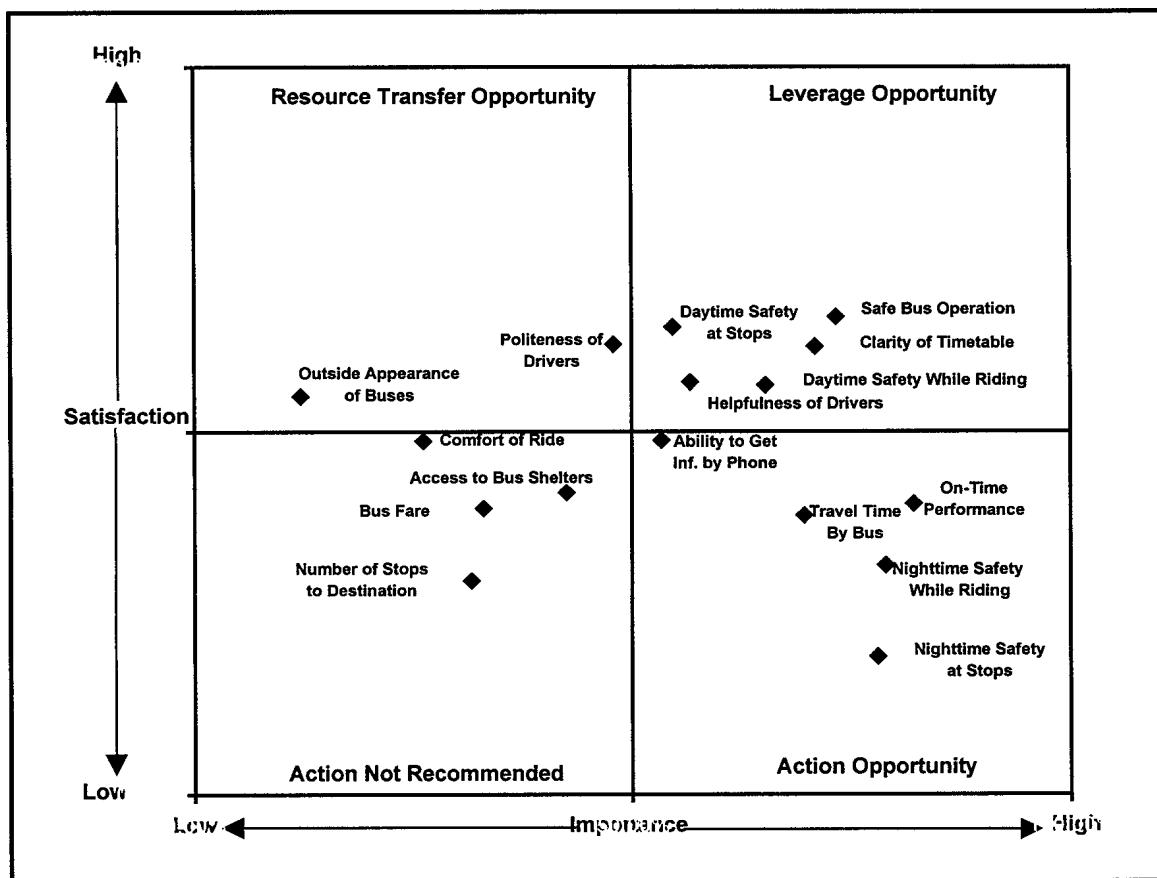
Based on this quadrant analysis, the recommendation would be to allocate resources to those areas that are important to customers and where satisfaction is low – nighttime safety while waiting for and riding the bus, travel time by bus, on-time performance, and ability to get information by phone – a very different picture than what was discovered by looking at importance and satisfaction scores separately.



## EXHIBIT 24 DATA FOR A TYPICAL QUADRANT ANALYSIS

	Importance (% Very Important)	Rating (% Very Satisfied)
On-Time Performance	82.2%	40.0%
Nighttime Safety While Riding	79.0	31.5
Nighttime Safety While Waiting	78.1	18.9
Safe Bus Operation	73.4	65.7
Clarity of Timetable Information	71.0	61.6
Travel Time by Bus	69.7	38.4
Daytime Safety While Riding	65.3	56.3
Helpfulness of Drivers	56.7	56.7
Daytime Safety While Waiting	54.7	64.3
Ability to Get Information by Phone	53.4	48.8
Politeness of Drivers	47.9	61.9
Access to Bus Shelter	42.5	41.6
Bus Fare	33.0	39.4
Number of Stops to Destination	31.6	29.4
Comfort of Ride	26.1	48.7
Outside Appearance of Buses	12.1	54.8

FIGURE 22  
SAMPLE QUADRANT ANALYSIS



## Reward-Penalty Analysis

Thus far, the analysis has looked at the distribution of a single variable and/or the relationships or associations between two variables. Several multivariate methods of analysis offer the opportunity to assess the relationships or patterns among more than two variables simultaneously. Some methods commonly used in customer satisfaction research include correlation analysis, multiple regression analysis, discriminant analysis, factor analysis, correspondence analysis, and conjoint analysis. Providing illustrations of all these methods is far beyond the scope and purpose of this handbook. However, to illustrate the strength and richness of these methods, one example is provided.

A frequent question that decision-makers want addressed is where to allocate resources. While the quadrant analysis illustrated above provides some insight into this question, it does not suggest where improvements are likely to reap the greatest benefits. A "reward – penalty" analysis – using multiple regression analysis – provides further insight into the impact of service improvements in specific areas on overall customer satisfaction.

Multiple regression is a statistical technique used to develop an equation that predicts or estimates the value of a dependent variable based on the values of one or more independent variables. Here, the equation is used to understand the impact of improvements in customers' ratings of individual performance on their overall satisfaction with the agency. Regression analysis helps in understanding the importance and predictive accuracy of the performance attributes. Multiple regression analyses are easy to generate with available computing technology. Unfortunately, however, the results of a multiple regression analysis may be difficult to interpret without the help of a professional statistician. Again, it is beyond the scope of this handbook to go into the details of how to do regression. Rather, the purpose of the following example is to give the reader an idea of the type of insight this type of analysis can lend to decision-making. If at some future date, regression analysis seems an appropriate method, the researcher should take extra care in ensuring an understanding of the underlying assumptions and theory inherent in the analysis.

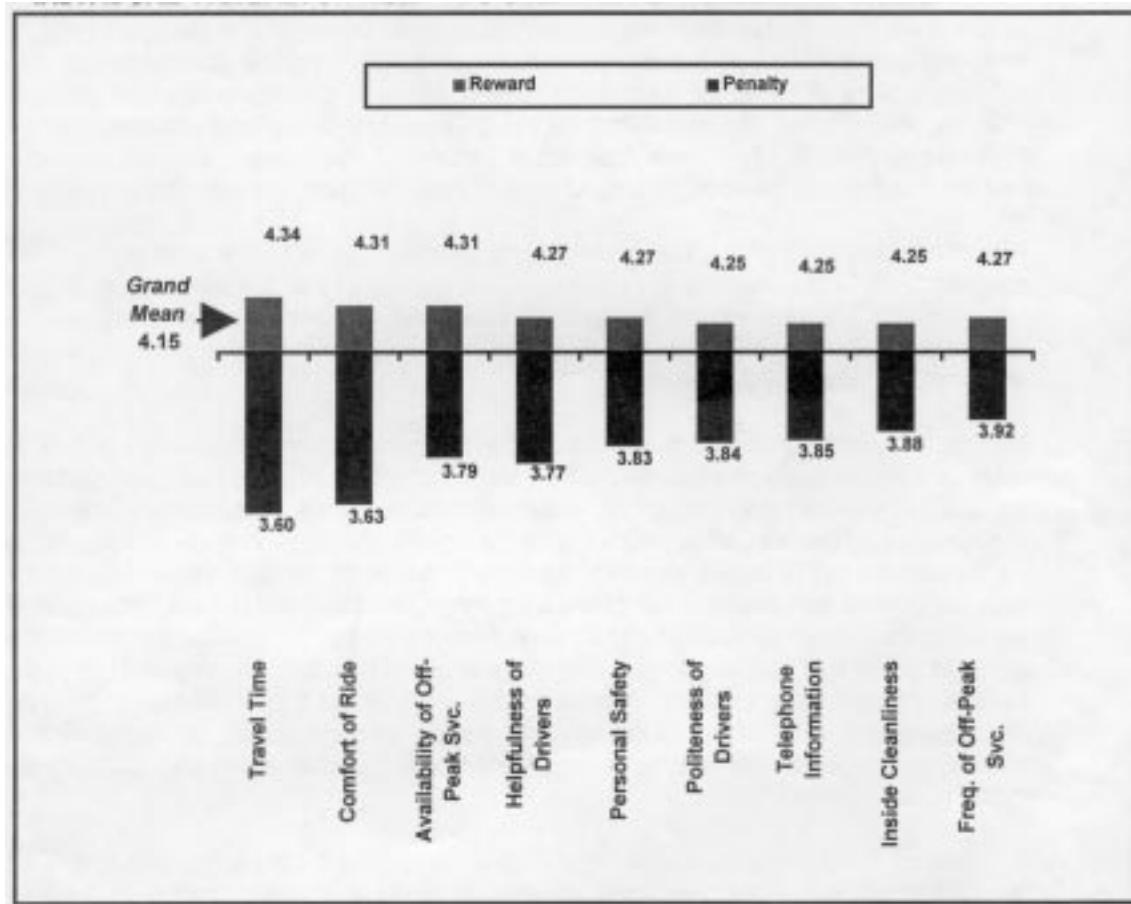
To accomplish this analysis, respondents were asked to rate a transit agency's performance as to its overall performance – the dependent variable – and as to its performance on twenty-three critical performance attributes – the independent variables. The resulting equation showed that nine of the twenty-three original independent variables explained over half the variation in the value of the dependent variable – overall performance. The equation then can be used to estimate the change that would occur in overall performance if the agency improves service in any single area. Conversely, the equation can be used to estimate the change that would occur in overall performance if quality of service declines in any one area. In our example, therefore, we can identify what will happen to the agency's overall performance score if all its efforts are placed in improving service in one area.

Figure 23 illustrates the results from this analysis. It shows that overall satisfaction will increase from 4.15 to 4.34 if the agency places all of its efforts into improving travel time. In other words, scoring a "5" on travel time, while all other issues remain constant, would improve the overall grand mean from 4.15 to 4.34. Conversely, the penalty numbers reflect the effect if the agency were to fail – receive a rating of "1" – on the specific criteria.

The analysis suggests that improvements in travel time are likely to achieve the greatest improvement in the overall performance rating. However, this improvement is not much greater than that which the agency would achieve if they made improvements in the other eight areas. On the other hand, decreases in travel time and comfort of the ride are likely to have the greatest negative impact on the overall performance rating. The recommendation would be to maintain and/or improve service in these two areas.



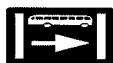
**FIGURE 23**  
**REWARD – PENALTY ANALYSIS – EFFECT OF CHANGES IN DELIVERY OF INDIVIDUAL CRITERIA ON AGENCY'S OVERALL PERFORMANCE RATING**



### Putting It All Together: High Quality, Actionable Research

As shown, when using the right tools, customer research – notably customer satisfaction research – creates a wealth of strategic information about customers, the transit agency, and the environment in which the agency operates.

The following Roadmap summarizes some simple strategies to keep in mind to conduct high quality, actionable customer research.



## ROADMAP 22 CONDUCTING HIGH QUALITY, ACTIONABLE RESEARCH

<input checked="" type="checkbox"/> <b>Use the voice of the customer.</b>	Ensure that the evaluation criteria against which an agency's performance is measured are based on attributes and terminology used in the market place. The criteria should reflect all elements that are part of the customer's evaluation process and reflect the thought process of the customer. Internal agency terminology should not be used during the interviewing process.
<input checked="" type="checkbox"/> <b>Include quantitative and qualitative information.</b>	Quantitative data is critical to the overall measurement effort. Qualitative information is equally important. Upfront qualitative information is helpful in designing a customer-oriented research effort. Following a survey, such information lends insight into the quantitative measures, bringing dynamics and causal data to the analysis. Qualitative information gives the "why" and "why not," "how," and "please explain" to the data.
<input checked="" type="checkbox"/> <b>Pretest the questionnaire and sample design.</b>	Pretests allow researchers to modify and finalize the questionnaire and sample design to incorporate the voice of the customer. Results of pretest interviews ensure that the evaluation criteria are properly identified, criteria are defined using the words of the customers, research methodology is appropriate and technically correct, and questionnaire obtains reliable and valid responses.
<input checked="" type="checkbox"/> <b>Define performance targets.</b>	To develop and implement improvement strategies that address the concerns of the marketplace, specific quality targets for key evaluation factors should be sought. Query respondents about their definitions of ideal and acceptable levels of performance. This information will help develop actionable steps for improvement.
<input checked="" type="checkbox"/> <b>Use an unbiased sample.</b>	When designing the sample, once the population of interest has been defined and sampling frame identified, use a random selection process. This will ensure that any particular group of respondents does not bias results – for example frequent riders.
<input checked="" type="checkbox"/> <b>Use experienced researchers.</b>	To obtain specific and accurate performance targets, use experienced researchers. They should be experienced in both customer research methods and in the industry under study. Experienced researchers are able to probe in depth as well as accurately and quickly assess respondent input. Moreover, experienced researchers will use the appropriate analytical tools.
<input checked="" type="checkbox"/> <b>Ask for improvement suggestions.</b>	Include questions that suggest information about specific areas of needed improvement, including product and service improvement areas. Seek suggested priority of improvement strategies. Incorporate these questions into the latter part of a questionnaire.
<input checked="" type="checkbox"/> <b>Include market-driven questions.</b>	Include questions on respondents' likelihood to continue to use transit or other measures of customer loyalty. These market-driven questions lend insight into the overall satisfaction of respondents.
<input checked="" type="checkbox"/> <b>Use techniques that emphasize improvement.</b>	When analyzing data, the most useful techniques are those that assist in identifying and prioritizing improvement requirements. Simple scorecard techniques are just that – simple. They provide minimal dynamics and strategic input. Use multivariate methods that not only show the priority of improvement strategies but also the impact of such strategies on overall performance.