

CUSTOMER SATISFACTION MEASUREMENT ISSUES IN THE FEDERAL GOVERNMENT

Tracy R. Wellens, Elizabeth A. Martin, U.S. Bureau of the Census

Tracy R. Wellens, SRD/CSMR, Room 3127, FOB 4, U.S. Bureau of the Census, Washington, D.C. 20233-9150

Key Words: Surveys, Benchmarking

Since the issuance of Executive Order 12862 for "Setting Customer Service Standards," customer satisfaction measurement has become prominent in the Federal Government. As part of "creating a government that works better and costs less" the National Performance Review suggested "putting customers first." The thrust of this initiative is to have the Federal Government function more like private industry. The competitive markets of the private sector have created a climate which is customer focused and the Federal Government is attempting to follow suit. The adoption of a customer orientation in the private sector is based on the assumption that satisfied customers will remain loyal to the provider. In the Federal Government this customer focus is more complex because the issue is politicized and not based solely on monetary gain.

The Executive Order for "Setting Customer Service Standards," requires Agencies to engage in several activities. The following activities from the Executive Order are all survey related:

a) identify the customers who are or should be served by the agency, b) survey customers to determine the kind and quality of services they want and their level of satisfaction with existing services and c) benchmark customer service performance against the best in the business which is defined as the highest quality of service delivered to customers by private organizations providing a comparable service.

In this paper we discuss several broad issues that face the Federal Government in its attempts to measure customer satisfaction. These issues will be discussed in the context of: questionnaire development, customer definition and sampling frames, response rates, confidentiality and data comparability.

These issues became relevant in our recent challenge of designing and conducting a department-wide customer satisfaction survey for the Department of Commerce. In this paper we provide a brief description of that survey and draw on that experience to illustrate many of the issues that face the Federal Government in the area of customer satisfaction measurement.

The Department of Commerce wanted one generic survey for the entire Department which covers 14 separate agencies. Some of these agencies are very large and for operational purposes are considered separate units. Taking this into consideration, the survey was expected to cover 20 separate operating units within the Department of Commerce.

This approach of creating one comprehensive generic survey attempted to standardize the measurement process

across the separate agencies and to minimize development costs by developing one product for use everywhere within the department. However, this approach led to two separate sponsors with two separate goals. The first sponsor who commissioned the developmental work was the Department of Commerce. The department wanted departmental-comparison information which could be used for decision making purposes, including budget allocations.

The second sponsor for this survey can be seen as the participating agencies who paid for the survey. The participating agencies wanted detailed agency-specific information. Each agency wants information to evaluate and hopefully improve their own customer satisfaction. As part of the attempt to solicit agency participation in the survey, the agencies were promised this very specific information.

The separate goals of these two types of sponsors are often competing in terms of the agencies need for specific information and the department's need for general comparison information. These two separate goals are seen in the survey instrument and contributed to its length.

There were several stages involved in the development of the questionnaire. First, we had to identify the types of products and services which were provided by the various agencies within the Department of Commerce. Unfortunately, there are no central lists of products and services so we had to generate them. It is also important to keep in mind that we were dealing with 14 agencies with very diverse aims and purposes, and subsequently diverse products and services. Products and services ran the gamut from Census data tapes, BXA export licenses, NWS weather forecasts, NOAA fishery inspections and disaster relief services, to ITA training seminars, etc.

After we compiled the product and service lists, we developed categories and began writing survey questions to target those categories. We were able to group all products and services provided by the DOC into three broad categories:

1. Information Services and Data Products

included: informational materials, such as newsletters, catalogs, promotional brochures, videos, telephone calls and personal visits, information fax lines, electronic bulletin boards, referral services, tours, informational reports and radio programs off-the-shelf data products and software.

2. Specialized Services or Products

included: customized services or products developed for specific organizations such as data collection, research, technical assistance, consulting, specially prepared tabulations, policy or negotiation services, disaster relief, standard reference materials, and training courses.

3. Grants and Regulatory Products and Services

included: grants, awards, licenses, certifications, accreditation, inspections, patents, trademarks.

These categories provided the framework for the survey. We developed 3 modules consisting of questions targeted to each product and service category. Although the products and services within a category were diverse, the types of questions asked about them were similar. Each agency's questionnaire only included those modules which were appropriate to the categories of products and services it offered.

In terms of questionnaire content, we had to determine the types of questions that would be applicable across the three product and service categories. We also needed to keep in mind the goals of our two sponsors. We decided to ask questions about all of the aspects involved in the process of obtaining and using products and services. We targeted such areas as: timeliness of the information, quality of the product or service, documentation, clarity, ease of use and price. We also asked questions about agency staff in terms of their competence, responsiveness, and handling of problems.

We used a combination of different types of questions when preparing the survey. We used rating scales, other close-ended and open-ended question formats. However, the survey consisted primarily of rating scales. These were used to measure specific levels of satisfaction with all aspects of product and service use, as well as to measure global, overall product/service satisfaction within each of the three categories. Rating scales were also used to assess how important each dimension of the product and service use was to the respondent.

We also included a global question for each section which asked if the products and services in this category met the respondents requirements and a "Bureaucratic Red Tape Index" was used to assess the amount of bureaucratic red tape which was necessary to obtain each category of products or services.

There are many factors that determine whether an evaluative survey is designed well. There are several issues related directly to the use of rating scales. These issues were discussed at length by Schwarz, (1995). However, assuming that the instrument is well designed and able to measure what was intended, there are still several issues to be resolved before the survey can be conducted.

First among these is who to survey. Determining who your customers are and what types of records are available for sampling are issues every customer survey has to resolve.

The executive order defines a customer as "an individual or entity who is directly served by a department or agency." At first this may seem straightforward, but this definition was debated for quite sometime. Some people interpreted this to include both internal and external customers. Internal customers are defined as those within the organization and external is defined as those outside the organization. For the Commerce Customer Survey, we decided to limit participation to customers outside of the Commerce Department. Since we knew that the department planned to compare results across the 20 operating units we decided that agencies should not survey their own employees nor employees of any other Commerce Department agencies who may be their customers. We also favored an interpretation of the Executive Order definition which focused on external customers. This is not to suggest that agencies should not also obtain information and feedback from internal customers. However, the types of questions you would ask and the formats for those discussions may be quite different from those for external customers.

Even after limiting this discussion to external customers, determining who your customers are is not always straightforward in the Federal Government. To most of us, a customer is someone who purchases a product or service, usually by choice or voluntarily. In a government setting, many products and services are not purchased directly by their users, but subsidized in whole or in part by taxes. Many government products and services are not received voluntarily on the part of the "user" or recipient. Many "customers" of enforcement services or tax collection services no doubt would, if they had the choice, choose not to obtain the service at all.

There are several types of customers in the Federal Government, ranging from the obvious to the not-so-obvious. Defining an agency's customers is likely to spur controversy. Here are some possible types of customers:

1) Customers who purchase products and services.

This is probably the most obvious and straightforward type of customer. Paying customers are those for which an agency is most likely to have good records, making them easiest to identify and therefore survey.

2) Customers who request and receive products and services for free.

This type of customer is not as straightforward. It has been argued by some that if people don't purchase things from you, why should they be considered a customer? One philosophical answer would be that as a Governmental agency, it is our duty to consider them to be customers. After all, someone (the taxpayer) is paying the agency to provide these products and services.

If that is not convincing enough, one should also consider that customers who receive products and services from an agency for free may be strong candidates to be paying customers in the future. An organization may want to survey these "potential" customers to find out if they would pay for products and services in the future. This is exactly

what we did in the Commerce Customer Satisfaction Survey.

The records an agency keeps for customers receiving free products or services may not be as good as the records they maintain for paying customers. If this is the case, it may be more difficult for an agency to survey this type of customer.

3) Customers who passively or even unknowingly receives products and services.

Similar to the last example, this type of customer may not pay an agency directly, but through taxes, may be paying for an agency's products and services. An example of a passive customer would be someone who listens to the radio to hear information from the National Weather Service. The National Weather Service has no record of who is listening, but this listener is clearly a customer who may or may not be satisfied with the service she is receiving.

Another difficulty arises because many services offered by the government are not intended to benefit those who experience them directly, but to protect or benefit others, such as the public, who may not even be aware of their existence. For example, one service provided by a Department of Commerce agency is the inspection of fisheries. Presumably this service is ultimately intended to benefit fish-eaters by ensuring the quality of fish, but these "customers" may not even be aware of the service.

Passive or unknowing recipients of services are probably the most difficult type of customer to survey. If the customer is not actively attempting to receive your products or services, they may not even know that your agency is providing them. Chances are, the agency does not know their identity either.

4) Customers who are regulated by an agency.

This customer may or may not pay for an agency's services. Because of the relationship the agency has with this type of person or organization, they may not even consider themselves to be an agency customer, especially since they may be obtaining the "service" involuntarily. But if an agency regulates this person's or organization's activity, they are probably providing a service. It may be worthwhile to find out if the person receiving the service is satisfied with the way the agency is providing it.

An example here might be a fishery that is regulated by NOAA. The process used by NOAA to regulate and inspect fisheries could be evaluated by surveying the fisheries.

As far as the ability to survey such "customers", chances are, if an agency regulates someone, they probably have good records of who they are and could therefore, survey them. The agency needs to decide if it would be appropriate to do so.

5) Congress as a customer.

There was much debate in the Department of Commerce about how Congress should be handled as a customer. One side was concerned that if you asked particular Congressmen how satisfied they were and then did not do intensive enough follow-up, negative repercussions would result. We decided to include Congress in our sample and

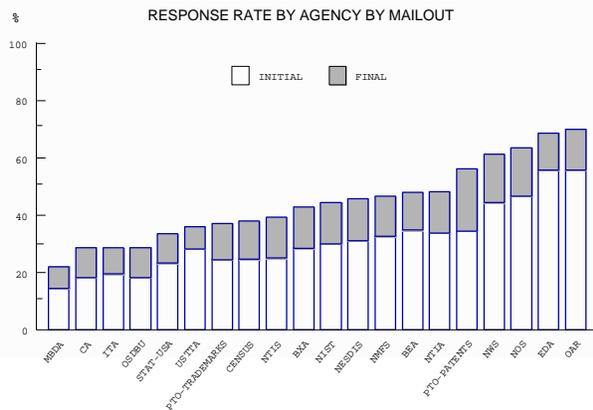
suggested that additional efforts could be undertaken to address the issues faced by Congressional customers more directly.

Due to time constraints, for the Department of Commerce Customer Survey, we decided on a general request to the agencies for obtaining customer lists. Each agency was instructed to provide lists of all customers who were external to the Department of Commerce. Unfortunately, this resulted in lists varying in quality and scope.

The customer lists delivered to the Census Bureau ranged in size from 114 to 190,000. This resulted in selected sample sizes ranging from 114 to 1500. The Census Bureau mailed out 21,970 questionnaires to customers of the 20 individual operating units within the DOC. (For a more detailed discussion of customer definition and sampling frame issues for the Commerce Survey see Ott and Vitrano (1995).)

The next issue to address is response rates. Unfortunately, response rates in the customer survey literature thus far have been inconsistent. With the exception of a few outliers, customer surveys achieving response rates above 30% are viewed by some as successful. Poorly designed questionnaires and survey implementation procedures have contributed to this low response rate. Several experts (who taught University of Maryland short courses on this subject) had proposed that customer surveys conducted by government agencies would have higher response rates than those in the private sector because they would be perceived as more credible. We believe that the opposite argument could also be made because of anti-government sentiments and resentment of more government "paperwork." Indeed, we know from interviews conducted with both respondents and nonrespondents in a pretest that some customers did not respond to the survey because they were mad at the agency. Based on discussions with OMB and other Federal Agencies, it is fair to say that response rates to government customer satisfaction surveys vary and that simply because the survey is conducted by the government will not necessarily lead to enhanced response rates.

Therefore, we wanted to be conservative in our estimates of the response rates we thought we could obtain. Nonetheless, we thought that we could maximize response rates through the use of a user-friendly questionnaire design and by incorporating a more comprehensive mailout procedure. We used an initial questionnaire mailout, a reminder card, and a second mailout of a replacement questionnaire to non-respondents. Research conducted by Don Dillman at the Census Bureau suggests that this procedure should increase response rates by at least 10 percentage points. (For more information on mailout procedures and response rates see Dillman, 1978.) Many surveys in the literature did not use any follow-up procedures. We also included a cover letter signed by an official whose name we hoped would be recognizable, Ron Brown, Secretary of Commerce.



As can be seen on the above graph, as expected our procedures did help response rates somewhat. Our overall response rate across all 20 operating units was 42%. Before the second mailout, the overall response rate was only 29.4% across all operating units. Thus, the mailing of a second questionnaire gained approximately 13 points overall. Although our overall response rate reached 42%, response rates across agencies ranged from 22% to 70%. Needless to say, these rates are not high. The decision to do more comprehensive follow-up has to be evaluated in terms of the quality of the additional information obtained and the costs involved in obtaining it. It should also be noted that collecting that information in person or over the telephone may result in mode effects, affecting data comparability. (See Hippler, Sudman and Schwarz, 1987; Schwarz and Sudman, 1992 for a discussion of response scale mode effects.)

In addition, another important factor related to response rates is the issue of confidentiality. Most people do not really distinguish between the concepts of confidential and anonymous. When data is confidential it means that you are linked to your response but you are being protected. An anonymous response cannot be linked back to the source.

The main issue then in is in the feeling of anonymity. If respondents feel that their responses are anonymous, in theory they are more likely to respond truthfully. This is very important with this type of satisfaction data -- particularly when the customer is involved in an ongoing relationship with the agency.

The use of a mail procedure can increase this feeling of anonymity. We intentionally avoided using identifying information on the form. We placed the address label on the envelope rather than on the questionnaire. That way the individual's name did not appear on the questionnaire form when returned. However, we were able to identify the individual questionnaires with a control number that was inconspicuously placed on the back cover. This was

necessary to determine who should receive a follow-up questionnaire.

It should be noted that the Office of Management and Budget (OMB) decided that this questionnaire did not fall under the privacy act so the Census Bureau could not legally assure the confidentiality of this data. We were able to and we did ensure that respondents' individual responses would be used only to produce statistical summaries. We at the Census Bureau also decided that none of the identifying information would be released to the agencies. How to handle the issues of confidentiality and identifying information will be a decision faced by federal agencies.

Returning to the purpose of conducting a customer satisfaction survey of this type, reminds us that the primary purposes are twofold. The first is to gain feedback about what an agency is doing right and wrong, and how it could be improved, and the second is to use the survey results as a benchmark of performance. The first of these purposes does not impose very strict requirements on how a survey is carried out and may not even require a formal survey. However, the second purpose, of benchmarking, imposes stricter requirements on data collection, because it implies comparison. Although there are several types of benchmarking which can be done (e.g., best practice, average performance, or baseline), all require that the data must be comparable across whatever units are to be compared.

Even with a well constructed scale for measurement, there are many other factors which need to be considered in determining whether a survey will or should be used to provide comparable data.

This paper will consider 3 types of benchmarks, all of which imply different comparisons. The first meaning of benchmark is as best practice or an ideal. The Executive Order clearly has this type of benchmark in mind when it states that "The standard of quality for services provided to the public shall be customer service equal to the best in the business." It specifies that agencies are to survey customers to determine their satisfaction with existing services, and benchmark their customer service performance against the best in the business in the private sector. This type of benchmarking implies comparing customer satisfaction between a government agency and a private company whose practice is identified as "best in the business." Thus, to use customer surveys to establish this type of benchmark, it is necessary that customer satisfaction be measured in a way that is comparable between the government agency and the targeted private company. This type of benchmarking also requires that best practice somehow be identified. That might be done outside of the customer survey itself, using external standards or criteria, or it might be part of the survey exercise, in which best practice is defined as that which elicits highest customer satisfaction. In the Department of Commerce Customer Survey, determining best practice is external to the survey itself. In order to this to be accomplished, at some later point, external criteria will

need to be located and compared. It is unclear whether this information is in existence or will need to be established.

A second type of benchmark is what other agencies or companies do, or average performance. This type of benchmarking implies comparisons among agencies and/or private companies. The Executive Order also appears to have this type of benchmarking in mind, when it states that "each agency shall use [customer satisfaction] information in judging the performance of agency management and in making resource allocations." This type of comparison requires that the measurements of customer satisfaction be comparable among agencies or companies which serve as benchmarks for each other. Obviously, if the information is to be used to make decisions about allocation of resources, one would want to be very certain that the comparisons are meaningful and that differences in customer satisfaction between agencies are not artifacts of the way the data were collected. This is important for the Department of Commerce Customer Satisfaction survey because this is the type of comparison information the department would like to have for decisionmaking purposes.

The third meaning of benchmark is as a baseline, with subsequent measures compared against it to measure improvement. In this type of benchmarking, the standard of comparison is an agencies' own prior performance, and one measures change or improvement against that. This type of benchmarking requires that repeated surveys be comparable to one another in order to make legitimate comparisons between the baseline measure of customer satisfaction and subsequent measures of satisfaction. This type of benchmarking is also implied by the Executive Order, for which agencies are expected to conduct repeated customer surveys and use the results to judge employee performance.

All 3 types of benchmarking then require comparisons among agencies or companies, over time, or both. There are several questions that need to be addressed in evaluating whether a customer survey can provide valid measurements to be used for benchmarking.

The first of these issues affecting comparability of data has to do with the identification and sampling of customers. In order to make comparisons across industries, one must be certain that the samples are comparable. In addition, one would want reasonably high response rates for all the agencies being compared. If response rates varied among agencies, then artificial differences in satisfaction may result from greater nonresponse bias for some than others.

In general, the construction of sampling frames for customer surveys is problematic. In this survey, samples were drawn from lists of customers provided directly by the 20 operating units of the Commerce Department. They were provided with instructions to try to ensure that customers were identified in a consistent way by different agencies. All agencies were instructed to include all customers who were external to the agency and the Department of Commerce on their lists. Although we have not yet fully examined the quality or completeness of the

lists provided, there is anecdotal evidence suggesting that they vary widely in quality. For example, one agency had almost 32,000 cases from the original list of 46,500 deleted before sampling because of duplication. Once customer was represented on that list 900 times.

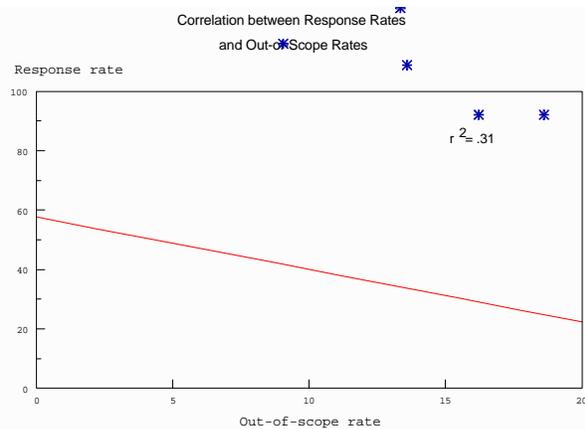
The very fact that the size of the lists varied from one list of only 114 customers, up to a list of 190,000 customers, strongly suggests that agency lists varied in completeness, probably in what they considered to be a "customer," and perhaps in the quality of their available records. We know that several agencies chose to exclude all non-paying customers who received promotional or other informational materials.

In addition to variation in quality, agency-provided lists are potentially vulnerable to selection bias, since organizational representatives who know that customer satisfaction is to be evaluated may overrepresent satisfied customers in their lists. In this survey several agencies acknowledged that they were only providing a sample of their actual customer base. We do not know of any cases in which an agency intentionally overrepresented satisfied customers but it is a potentially serious threat to comparability which must be kept in mind.

A more measurable threat to the comparability of satisfaction measures across agencies in this survey is nonresponse bias, arising from generally low response rates, and the very large differences in response rates for different agencies. As mentioned earlier, response rates varied from a low of 22 percent to a high of 70 percent.

The differential quality of the lists partly explains these differences. For example, the customer list of the agency with the lowest response rate did not identify individuals by name, but only listed "owner" for each organization and company the agency had served. Thus, questionnaires could not be mailed to specific individuals, and probably a large number were lost or misplaced. In addition, this agency only provided customer information for fiscal year 1993.

We examined differential list quality in terms of response rates and out-of-scope rates by each agency. Out-of-scope rates are defined as post-master returns without address corrections and other circumstances that were deemed out-of-scope (i.e., respondent no longer employed at organization or company out of business). The out-of-scope cases are taken out of the base N before a response rate is calculated.



Looking at the plot above, a relationship emerges with an $r^2 = .31$. Thus, those agencies with lower response rates had higher out-of-scope rates. This suggests that those agencies with low response rates had lists which did not represent current or recent customers. In terms of the agency with the lowest response rate, we know that the agency only provided customer information for fiscal year 1993. Moreover, the customer list did not identify individuals by name, but only listed "owner" for each organization and company the agency had served. Thus, questionnaires could not be mailed to specific individuals, and probably a large number were lost or misplaced. It is understandable why these low quality customer lists would contribute to low response rates.

Another factor likely contributing to differential response rates is self-selection. Respondents are likely to represent customers with stronger feelings--either positive or negative--compared to those who do not take the time to respond. When the response rates are very low, then self-selection bias is likely to be severe, although its effect on comparisons of mean satisfaction levels across agencies may not be obvious.

In summary, customer surveys which aim to compare across diverse agencies and products potentially are affected by very serious problems of data comparability, including lack of comparability arising from sample design, differential nonresponse, and the measurements themselves. There appears to be a considerable amount of careful methodological and statistical work that still needs to be done to ensure that customer surveys are designed to yield meaningful comparisons of customer satisfaction across industries and over time. Until that groundwork is done, such comparisons should be made very cautiously. (For more details on this topic and result highlights from this survey see Wellens, Vitrano, & Martin, (1995).)

In addition to the careful attention to issues related to data comparability, there are a couple of other key points that need to be made about these surveys. First of all, there needs to be clear specification of the goals of the survey, and an understanding of how the information from the survey will be used. This is very important in a political

climate. Second, customer surveys can themselves affect the expectations of customers and employees. Carrying out a customer survey may raise the expectations of customers (and employees) that a company or agency is going to do something to improve service. If that doesn't happen, and if the survey turns out to be an empty exercise, then the indirect effect of a customer survey may be to reinforce the cynicism of customers or employees or both.

Discussion of these issues is very important given the track record for customer surveys in the literature thus far and the expansion of their use in the Federal Government. Given the costs involved in designing and conducting these surveys and the importance of the decisions which may be based on them (e.g., budget allocations and employee performance appraisals), it is essential to understand the uses and explore the limitations of this type of measurement.

References

- Dillman, D. (1978). Mail and telephone surveys, the total design method. New York: Wiley and Sons, Inc.
- Hippler, H. J., Schwarz, N., & Sudman, S. (Eds.). (1987). Social Information Processing and survey methodology. New York: Springer-Verlag.
- Ott, K. & Vitrano, F. (1995). Frame development and sample selection issues for the Commerce department-wide customer satisfaction survey. Paper presented at the annual meeting of the American Statistical Association, Lake Buena Vista, FL.
- Schwarz, N. (1995). How do respondents answer satisfaction questions? Paper presented at the annual meeting of the American Association for Public Opinion Research, Ft. Lauderdale, FL.
- Schwarz, N. & Sudman, S. (Eds.) (1992). Context effects in social and psychological research. New York: Springer-Verlag.
- Wellens, T., Vitrano, F. & Martin, E. (1995). Questionnaire development, results and other issues from the Commerce department-wide customer satisfaction survey. Paper presented at the annual meeting of the American Statistical Association, Lake Buena Vista, FL.