

# RESULTS FROM A CUSTOMER SERVICE SATISFACTION MAIL SURVEY USING DILLMAN'S "TOTAL DESIGN METHOD" TO ACHIEVE HIGH RESPONSE RATES

Leonard V. Covello, U. S. Department of Agriculture  
Farm Service Agency, ITSD, P.O. Box 2415, Washington D.C. 20013

**KEY WORDS:** TDM, Question Order, Incentive

## Introduction

Those of us interested in assessing customer satisfaction may be engaged in a new version of Lewis Carroll's Snark hunt. "Hunting for Snarkes is a very pleasant occupation if you do but make-believe strong enough" (*Temple Bar*, 1879. *Compact Edition of the Oxford English Dictionary Vol II*, 1971, p. 309). We may disagree about many qualities of our Snark but most of us seem to agree that we can catch it with a survey. Confounding the theoretical or conceptual hunt for satisfaction qua Snark is a methodological fact of life, all types of sample surveys are subject to various types of errors: sampling error, noncoverage error, nonresponse error and measurement error (Groves 1989, Dillman 1991). Two of these error sources, sampling and noncoverage error feel as if they can be controlled conceptually by sample size/confidence intervals and a good sampling frame. Measurement and nonresponse error seem less controllable.

Two aspects of measurement error contribute to satisfaction's Snark-like, elusive quality. One aspect is conceptual and the other is methodological. Far from resolved are the questions, what is satisfaction (conceptual) and how is it measured (methodological). Yi, for example, identified multiple studies variously conceptualizing satisfaction as an *outcome* resulting from a consumption experience while others conceptualize it as an *evaluation* of a consumption process (Yi 1990). Methodologically, single and multi-item measures as well as unipolar and bipolar scales have been used. The number of response categories has also varied from two to very large numbers (Yi 1990, Peterson and Wilson 1992, Devlin, Dong and Brown 1994). One result of these measurement weaknesses is the consistent finding that surveys report distributions characterized by large positive bias and negative skew (Westbrook 1980, Peterson and Wilson 1992, Devlin, Dong and Brown 1994).

But settling on a single conceptualization and a measurement approach is not enough. Even when conceptualization and measurement are resolved nonresponse error enters to reconfound. Presumably, the satisfaction survey is conducted because the *actual*

distribution of satisfaction in the target population is *unknown*. Because it is unknown, nonresponse leaves the investigator with the unsatisfactory task of estimating an *unknown* attribute about an *unknown* target—the nonrespondents (Dillman 1991). While it might be true that the distribution of the surveyed attribute (satisfaction in this case) is the same among respondents and nonrespondents alike, that conclusion usually requires an act of faith resting on similar demographic characteristics of the two groups. Because of this problem, "...response rate has de facto become the accepted proxy for nonresponse error. The higher the response rate the lower the potential of nonresponse error and therefore the better the survey" (Dillman 1991, p. 229).

This paper explores the issues of nonresponse and measurement error in one national mail survey of customer satisfaction. This survey attempted to reduce nonresponse error by increasing response rate and sought to reduce measurement error by reducing the usual positive bias of responses and producing some reasonable variability.

**Nonresponse Error** To increase response rates, the survey primarily relied on a particular implementation of "social exchange theory" called Total Design Method (TDM). TDM was developed by Don A. Dillman of Washington State University (Dillman 1978). Dillman reports that the TDM approach achieves response rates comparable to face to face interviews. (Dillman 1978, 1993). Recently, Dillman published the results of two research efforts designed to measure the independent contributions on response rates of five TDM techniques: respondent friendly construction (3.4 percent), slightly shorter form (4.6 percent), replacement questionnaire (10.4 percent), prenotice letter (6.3 percent). The estimated response rate from these techniques was 71.4 percent versus 40.3 percent on a previous survey not using these techniques. This paper provides a comparison of results using some of these techniques, differentially, in the pilot and the full survey implementation. In this sense, it provides additional evidence to assess the utility of TDM.

In a meta-analysis of incentives on mail survey response rates, Church reported that surveys using a non-monetary incentive achieved a 7.9 percent average

improvement in response rate over controls (Church 1993). Although Dillman, in the studies reported above, did not measure the effect of an incentive, TDM provides for them. "Social exchange" implicitly assumes that an incentive is an inducement to cooperate. The use of a nonmonetary incentive was tested in the satisfaction pilot reported in this paper. Pilot results are compared to Church's report.

**Measurement Errors** Several aspects of measurement error associated with satisfaction are explored here. First, the survey contains some data about the issue of part-whole question order (Shuman and Presser 1981, Schul and Schiff 1993, Mason, Carlson and Tourangeau 1994). The effects of using unipolar 5 and 6 category scales on the positive bias/negative skew phenomenon noted by Westbrook (1980) and Peterson and Wilson (1992) are also reported.

This paper also describes the ability of respondents to make useful estimates of "waiting times" and the effect of this wait on customer service and satisfaction. This information is reviewed in terms of Taylor's contention that perceptions of delays are an important determinant of assessments of quality service and that such estimates have empirical validity because they correlate with actual delays (Taylor 1994). Survey data about delays were compared to measurements of actual delays to determine the accuracy.

Finally, the issue of estimating satisfaction in the population, given nonrespondents, is also addressed. Some have hypothesized that reported satisfaction overestimates population satisfaction because respondents are likely to be more satisfied than nonrespondents. Peterson and Wilson reported that they found no meaningful correlation between satisfaction percentage and response rates ( $r = .05$  in one study and  $.06$  in a second). This approach only addresses the percentage of satisfied *respondents*. Dissatisfied *nonrespondents* were not included. Since TDM relies on multiple mailings or contacts, it is possible to see if reported satisfaction varies by when someone responds. In other words, early respondents may be thought of as cooperative and nonresistant. The last respondents may be less cooperative and more resistant. Compared to early respondents, they might resemble the least cooperative and most resistant part of the sample, *i.e.* those who do not respond at all. This paper also reports the results of the relationship of satisfaction and response rate, using this conceptualization.

## Method

### Data Source

Data reported in this paper are taken from a survey and pilot survey of customer satisfaction conducted by an agency of the U. S. Department of Agriculture.

When conducted, the organization was called the Agricultural Stabilization and Conservation Service (ASCS). It is now called the Farm Service Agency (FSA). ASCS/FSA administers programs designed to improve and protect farm income. In simplified terms, it gives farmers payment checks in exchange for their agreement to help maintain farm prices by reducing production and preserving fragile lands through conservation practices and land retirement. In addition, it administers disaster assistance and farm loan programs.

The target population was individual farmers who visit the 2500 plus ASCS/FSA field offices to transact business. A pilot was conducted from June to July 1994. The full survey was started in September, 1994. Both the pilot and the survey used simple random samples, proportional to size and stratified by physiographic regions. The sampling procedure tended to insure representation by type of farming enterprise and production system. The frame was the organization records of payments recipients in 1993—the 1994 list would not be available until 1995.

### Total Design Method (TDM)

TDM uses social exchange theory to guide and integrate specific techniques and procedures. This theory, "...posits that questionnaire recipients are most likely to respond if they expect that the perceived benefits...will outweigh the perceived costs..." (Dillman 1991, p. 233). Some of the techniques and procedures Dillman has developed include: 1. Prenotice letter, using official stationery, explaining the purpose and usefulness of the survey, to the respondent and how results will be used; 2. User-friendly questionnaire with a cover letter, again explaining the importance of the survey and the importance of responding; 3. Three additional, carefully spaced mailings—a follow-up postcard and two separate replacement questionnaires and cover letters to nonrespondents. After the prenotice letter, all mailings contain prepaid return envelopes. Personalization is stressed throughout to make the materials stand out from junk mail. The letters are individually addressed. The envelopes are individually stamped and the letters are individually signed (or made to appear so). The details of design and administration are well documented (Dillman 1978).

The ASCS/FSA satisfaction survey and pilot used TDM techniques. The major deviations from classic TDM were that the letters were not individually signed, business reply rather than individually stamped *return* envelopes were used and the final mailing was ordinary first class mail not special delivery. The TDM techniques used in the pilot and full survey differed in that the pilot tested a nonmonetary incentive and did not employ a second follow-up mailing. All other treatments were identical. Table

1 summarizes the treatments.

**Table 1. TDM Elements**

TDM Treatment	Time	Pilot	Survey
User Friendly Design		YES	YES
Prenotice Letter	Week: Minus 1	YES	YES
Questionnaire:			
• Letter			
• Prepaid, Return Envelope	Week: Zero	YES	YES
• Incentive		Split Sample	NO
Postcard	Week: Plus 1	YES	YES
1 <sup>st</sup> Replacement Questionnaire	Week: Plus 3	NO	YES
2 <sup>nd</sup> Replacement Questionnaire	Week: Plus 7	NO	YES

In his research at U. S. Bureau of the Census, Dillman, *et al* tested the independent effect of various treatments and found the following independent contributions on response rates: respondent friendly construction (3.4 percent), slightly shorter form (4.6 percent), replacement questionnaire (10.4 percent), prenotice letter (6.3 percent). Because of the large sample used, individually signed letters was not tested. An incentive was also not used.

The ASCS/FSA sample matches the Census tests on four of the five treatments: respondent friendly construction, replacement questionnaire and prenotice letter. Both the pilot and survey used the follow-up postcard as did Census. However, the ASCS/FSA survey also used a second replacement questionnaire as described in classic TDM. The pilot and survey used a longer questionnaire than was used in the Census test. Deleting the effect of the shorter questionnaire matches the treatments used in the Census test with the ASCS/FSA pilot and survey.

**Response Rates** The results of the comparison are reported in Table 2. Although the return rates are significantly different ( $\alpha = .05$ , 1.6 SE<sub>p</sub>), the difference is a function of the large sample size. The size of the Census sample (50,000) virtually assures statistical significance, *i.e.* the ability to assert that observed differences are not likely to be chance phenomena. Statistical significance aside, these differences between the reported rates are not meaningful. If anything, these results seem to support the ability of TDM to help produce acceptable return rates for mail surveys.

**Incentive** The pilot survey included as its incentive a green and white bumper sticker that read, "I Gave [name & logo] a Piece of My Mind". Various versions of this bumper sticker were informally tested on farmers before testing in the pilot. The bumper

**Table 2. Response Rates--Percent**

TDM Treatment	Pilot Percent	Survey Percent	Dillman Percent
No TDM			40.3
Basic Treatment			
• Prenotice Letter			
• Questionnaire	51.5	57.7	56.4
• Postcard			
1 <sup>st</sup> Replacement Questionnaire	NA	65.2	66.8
2 <sup>nd</sup> Replacement Questionnaire		70.9	NA

sticker was distributed with the prenotice letter to a randomly assigned, split-half ballot. The letter contained an invitation to use the item so that friends and neighbors would identify them as survey participants.

In a meta-analysis of both monetary and nonmonetary incentives on mail survey response rates, Church reported that surveys using a non-monetary incentive achieved a 7.9 percent average improvement in response rate over controls (Church 1993). The types of nonmonetary incentives reviewed were diverse, *e.g.* coffee, pens, key rings, tie clips, etc. The results of the ASCS/FSA pilot are contained in Table 3.

**Table 3 Response Rate With and Without Incentive**

Pilot Survey	
With Incentive	Without Incentive
41.6%	51.5%

These results do not mean that nonmonetary incentives will necessarily *reduce* response rates. But it is clear that they will not automatically increase them. One plausible explanation was offered by Dillman. He suggested that since the population was generally satisfied (more than 90% expressed satisfaction on each of two separate measures) the message on the bumper sticker was dissonant with their affect toward the service provided. This explanation has some face validity. After all, why should pens, key chains, etc. increase rates and the bumper sticker not do so as well?

In a yet to be published study conducted by the National Customer Feedback Center, Veterans Health Administration, Department of Veterans Affairs, key chains, pens and *bumper stickers* were tested by split ballot. Those incentives showed no measurable effect on response rate. But they did not produce a reduction either. While Dillman's proposed explanation is appealing the effect produced in the ASCS/FSA pilot might be caused by other variables. These results point to the need for more *targeted*

research to help guide the selection of incentives that work.

**Satisfaction Question Order** The issue of question order, part-whole or specific-general question sequences, has been shown to produce somewhat contradictory results (Mason, Carlson and Tourangeau 1994). Schul and Schiff (1993) seek to identify conditions that create some of these contradictory results. They offered and tested three hypotheses about the effect of ordering general satisfaction and domain-specific questions (Schul and Schiff 1993, p 540). This ordering may be general satisfaction followed by domain-specific satisfaction (GS order) or the reverse (SG ordering). In their first hypothesis (H1) they first posit (H1.1) that in the GS order, *negative* responses to the general satisfaction question and *negative* responses to domain-specific questions will be more strongly correlated than the correlation between *positive* responses to the general satisfaction question and *positive* responses to domain specific questions. They expect this relationship because they believe that negative experiences are more likely to be remembered when the unprompted general satisfaction question is asked first. Therefore, in the GS order, prior negative experiences are more likely to be spontaneously activated initially. The SG order is less likely to show this pattern because the earlier domain-specific questions are expected to activate both positive and negative information. The second part (H1.2) of this hypothesis derives from the assumption that the respondent, free from the effect of relying heavily on prior negative experience, will recall both positive and negative experiences. This conceptualization also gives rise to the second hypothesis (H2). H2 posits that in SG order the correlation between domain-specific items and general satisfaction will be higher than in the GS order. In a third hypothesis (H3), they further posit that the *magnitude* of general satisfaction will be higher in the SG than in the GS order. These three hypotheses can be examined using the ASCS/FSA survey data.

When the ASCS/FSA survey was designed it was not clear if part-whole/specific-general order effects would occur and what those effects might be. For this reason and because two measures of general satisfaction were desired, the questionnaire was constructed with both GS and SG sequences. The initial question is a general, overall satisfaction item. It is followed immediately by domain-specific questions about satisfaction with the receipt of payment checks, questions about customer service in the office, satisfaction with forms used and satisfaction with participation rules. This sequence of items is GS order. The penultimate survey item is a question about general satisfaction with the overall quality of service. This question and the domain-specific questions which precede it, form an SG order. The results of operationalizing Schul and Schiff's

hypotheses are reported in Table 4. H1 predicts two conditions. The two conditions of H1 predict are:

$$H1.1: r_{GS\_Neg} > r_{GS\_Pos}$$

$$H1.2: |[ (r_{GS\_Neg}) - (r_{GS\_Pos}) ] | - |[ (r_{SG\_Neg}) - (r_{SG\_Pos}) ] | = 0$$

H2 holds:

$$H2: r_{SG} > r_{GS}$$

H3 changes from a measure of association to one of magnitude. It posits:

$$H3: P_{SG} > P_{GS}$$

Hypotheses H1 and H2 are examined using Fisher's r to z transformation. A weighted average correlation and z statistic were computed for each group of domain-specific items using Snedecor and Cochran's procedure (Snedecor and Cochran, 1967, p. 187). The correlations and their corresponding z values in Table 4 are not consistent with either H1.1 or H1.2. The correlations are the opposite of the prediction and are significant well beyond the  $p < .001$  level. H2 is also not supported. Although three of the four observed differences are the opposite of the prediction, none are large enough to be distinguishable, beyond chance, from no difference. The proportion of satisfied respondents in both the GS and the SG sequences are not significantly different. This result is therefore not consistent with the expectations expressed in H3 and those offered by Shuman and Presser.

**Table 4** Product Moment Correlations and Fisher r to z Transformations

Domain Specific Satisfaction	General Satisfaction			
	GS		SG	
	$\bar{r}$	$\bar{z}$	$\bar{r}$	$\bar{z}$
<b>Negative</b>				
● Payments	.070	.071	.187	.191
● Office Service	.226	.231	.207	.211
● Forms	.168	.172	.040	.041
● Eligibility Rules	.178	.183	.110	.108
<b>Positive</b>				
● Payments	.310	.318	.336	.348
● Office Service	.422	.449	.414	.437
● Forms	.254	.260	.236	.239
● Eligibility Rules	.264	.265	.291	.302
<b>All Responses</b>				
● Payments	.389	.406	.380	.397
● Office Service	.523	.583	.515	.565
● Forms	.300	.305	.319	.327
● Eligibility Rules	.327	.341	.345	.360

**GS: P=.916 SG: P=.919**  
**(SE<sub>p1</sub>, p<sub>2</sub>=.008 H3: NS)**

**Positive Bias and Negative Skew** Satisfaction surveys have consistently reported distributions characterized

by large positive bias and negative skew (Westbrook 1980, Peterson and Wilson 1992, Devlin, Dong and Brown 1994). Peterson and Wilson, for example, report that in Lebow's meta-analysis of 26 satisfaction studies, 12 percent of the studies report satisfaction percentages between 91 and 100 percent. In their own meta-analysis they report that the two highest categories average 82.9 percent of respondents. These findings show virtually no variance by which meaningful change could be observed over time. Tables 5 and 6 contain the distribution of satisfaction responses from the ASCS/FSA survey. These unipolar scales produce more variance than the typical bipolar polar scales.

**Table 5.** Distribution of General Satisfaction Responses—Percent

Satisfied	Overall Service	Overall Quality
	Percent	Percent
• Slightly Satisfied	2	2
• Somewhat Satisfied	7	5
• Pretty Satisfied	31	32
• Very Satisfied	46	41
• Extremely Satisfied	14	21
<b>Dissatisfied</b>		
• Slightly Dissatisfied	21	18
• Somewhat Dissatisfied	38	41
• Pretty Dissatisfied	28	23
• Very Dissatisfied	14	12
• Extremely Dissatisfied	0	7
<b>Neither</b>	5	

**Table 6.** Distribution of Domain Specific Satisfaction Responses - Payment Checks

Domain Specific	Percent
<b>Wait for Checks</b>	
• Not Short at All	18
• Slightly Short	16
• Somewhat Short	25
• Pretty Short	25
• Very Short	12
• Extremely Short	3
<b>Check When Expected</b>	
• Not at All Satisfied	5
• Slightly Satisfied	9
• Somewhat Satisfied	16
• Pretty Satisfied	31
• Very Satisfied	30
• Extremely Satisfied	9

**Respondent Estimates/Recollection of Time** When respondents are asked to recall and report on such matters as waiting time for service the usual assumption is that self reports of this time are inaccurate. ASCS/FSA respondents were asked to report on the length of their wait for payment checks. The reported times were longer than management believed was either reasonable or likely to have occurred. For that reason, the respondent reports were compared to the actual waiting times calculated from the computer generated, "payment history" archive files created during the application-payment process in each office. These results are reported in Table 7. A zero means the check was received on the same day as the application/approval occurred.

**Table 7** Reported and Actual Waiting Times for Payment Checks—Days

Payment Type 1	Min	Q1	Median	Q3	Max
Reported Wait	0	0	7	21	180
Actual Wait	0	7	14	27	184
<b>Payment Type 2</b>					
Reported Wait	0	0	14	30	203
Actual Wait	0	0	1	6	162
<b>Payment Type 3</b>					
Reported Wait	0	0	3	7	90
Actual Wait	0	0	2	6	307
<b>Payment Type 4</b>					
Reported Wait	0	0	1	10	240
Actual Wait	0	0	2	19	306

These data show that, at least under some circumstances, respondents can accurately recall and report information on wait time. At least, they do not tend to exaggerate the time. It is possible that these respondents were able to make these reports because the payments for which they were waiting, are particularly salient. Of course, it is also possible that they do not actually recall the time to the wait but rather have experience waiting about the same amount of time over the years. What they would then be reporting is a more general memory that recalls that the last time was like other times.

**Nonrespondents** In the ASCS/FSA survey about 29 percent of the eligible sample did not respond. If the nonrespondents were similar to some of the respondents, it would be possible to estimate their satisfaction level. TDM relies on multiple contacts, over time. Each TDM contact after the initial mailing and followup postcard, is directed to *nonrespondents*—i.e. a group that may simulate those who will never respond. The hypothesis is that the measure of self-reported satisfaction will differ by the timing of TDM contact. Specifically, those who respond later are presumed to have lower levels of satisfaction than those who respond earlier. These later responders are presumed to be the members of

the sample who are the least cooperative and most resistant to organizational entreaties to participate. These characteristics are further presumed to resemble the characteristics of the final nonrespondents. A GLM procedure was used to examine the relationship of TDM mailing and reported satisfaction. These results are presented in Table 8.

**Table 8** GLM of General Satisfaction Reports and TDM Mailing

Satisfaction by TDM Mailing	Mean	Pr>F Satisfied	Pr>Z
<b>General</b>			
<b>Satisfaction-Question 1</b>		.0001	.01
Initial Mailing	1.662	90.5	
1 <sup>st</sup> Replacement	1.508	93.3	
2 <sup>nd</sup> Replacement	1.477	85.1	
<b>General</b>			
<b>Satisfaction-Question 2</b>		.0001	.01
Initial Mailing	1.695	92.6	
1 <sup>st</sup> Followup	1.544	89.5	
2 <sup>nd</sup> Followup	1.508	85.1	

**Conclusions** There seems little doubt that TDM works. Also, TDM may help surveys reach portions of the population who are least satisfied. More research is needed to determine when incentives, especially nonmonetary incentives, will work. Question order may not be as critical as is sometimes believed. Respondents may in fact be able to accurately recall and report about service delays—at least if they are salient.

## References

- Converse, J.M., Stanley Presser. *Survey Questions: Handcrafting the Standardized Questionnaire*. Newbury Park, CA: Sage, 1986.
- Church, Allan H. "Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis", *Public Quarterly*. Vol 57, No1, Spring 1993, pp. 62-79.
- Devlin, Susan J., H.K. Dong and Marbue Brown. "Selecting Scales for Measuring Quality", *Marketing Research*, Vol. 5, No. 3, pp. 12-17.
- Dillman, Don A. *Mail and Telephone Surveys: The Total Design Method*. New York: John Wiley and Sons, 1978.
- Dillman, Don A. "Design and Administration of Mail Surveys", *Annual Review of Sociology*. 1991, pp. 225-49.
- Dillman, Don A. Kirsten K. West. "Influence of an Invitation to Answer by Telephone", *Public Opinion Quarterly*. Vol. 58, No. 4, Winter 1994, pp. 557-68.
- Dillman, Don A., Jon R. Clark, and Michael D. Sinclair. "How Prenotice letters, Stamped Envelopes and Reminder Postcards Affect Mailback Response Rates for Census Questionnaires". Paper presented at the Annual Research Conference of the U.S. Bureau of the Census, Alexandria, VA, March 21-24, 1993.
- Groves, Robert M. *Survey Errors and Survey Costs*. New York: John Wiley and Sons, 1989.
- Mason, Robert T., John E. Carlson and Roger Tourangeau. "Contrast Effects and Subtraction in Part-Whole Questions", *Public Opinion Quarterly*, Vol. 58, No. 4 Winter 1994, pp. 669-78.
- Peterson, Robert A. and William R. Wilson. "Measuring Customer Satisfaction: Fact or Artifact", *Journal of the Academy of Marketing Science*, Vol. 20, No 1, pp. 61-71.
- Schul, Yaacov and Miriam Schiff. "Measuring Satisfaction With Organizations: Predictions from Information Accessibility", *Public Opinion Quarterly*. Vol. 57, No 4, Winter 1993, pp. 536-50.
- Schuman, Howard, and Stanley Presser. *Questions and Answers to Attitude Questions: Experiments in Question Form, Wording, and Context*. New York: Academic Press, 1981.
- Snedecor, Gorge W. and William S. Cochran. *Statistical Methods* (6<sup>th</sup>Edition). Ames, IA: University of Iowa Press, 1967.
- Taylor, Shirley. "Waiting for Service: The Relationship Between Delays and Evaluations of Service", *Journal of Marketing*. Vol. 58, April 1994, pp. 56-69.
- Westbrook, Robert A. "A Rating Scale For Measuring Product/Service Satisfaction", *Journal of Marketing*. Vol. 44, Fall 1980, pp. 68-72.
- Youjiae Yi. "A Critical Review of Consumer Satisfaction", In *Review of Marketing*, Valarie A. Zeithaml ed., Chicago, IL: American Marketing Association, 1990, pp 68-123.