EFFECTS OF RESPONSE DELAY ON THE QUALITY OF SATISFACTION DATA IN MAIL SURVEYS

William R. Forrester, Jr., Kennesaw State College
Manfred F. Maute, McGill University

ABSTRACT

This study reports results of a test of the effects of response delay on four indicators of satisfaction data quality; item omission, section omission, response set bias, and response variability. Results support the study's major premise that the quality of satisfaction data decreases with response time. Late respondents to a mail survey of bank service customers had significantly higher rates of item and section omission than did early respondents. These findings underscore the importance of response speed in mail surveys and suggest that researcher reliance upon single indicators of response quality may lead to assessments of data quality which are incomplete and inaccurate.

INTRODUCTION

Data quality issues are receiving increasing attention in marketing and consumer behavior and are quickly replacing response rate issues as the central focus of mail survey methodology research (Hawkins, Coney, and Jackson, 1988). Unlike traditional studies, which focus on response rates, data quality studies focus on issues concerning the accuracy and usefulness of the data which are actually collected (Groves 1987; Goetz, Tyler, and Cook 1984. p. 151). Since much of the research on consumer satisfaction/ dissatisfaction and complaining behavior (CS/D&CB) relies on mail surveys for data collection, an understanding of the factors that affect data quality is of fundamental importance to CS/D&CB researchers.

Indicators of Data Quality

While "accuracy" would appear to be the most meaningful indictor of data quality, the study of data accuracy is limited by the need to compare a "true score" with a reported score. Consumer behavior research, including CS/D&CB research, is often concerned with attitudinal variables for which "true scores" are not available. Consequently, data usefulness, as measured by such indicators as item omission, response set bias, and response variability, are becoming more common and are likely to emerge as the primary data quality issues in CS/D&CB research (Rubens 1989, Groves 1987).

Data usefulness is measured by indicators such as data completeness and response variability which affect the strength of association among variables. The presence of response bias has also been used as a measure of data quality, as has the presence of "don't know", "no opinion", and "not applicable" responses, since their absence may signal spurious responses (Goldsmith 1988, 1989, Duncan and Stenbeck 1989, Durand and Lambert 1988, Goudy 1976).

Time Dependencies

To date, only limited attention has been given to the relationship between data quality and response delay. Although numerous studies have investigated response time, most have treated it as a dependent variable. See Dommeyer (1988, 1989) and Wunder and Wynn (1988) for a discussion of the factors which affect response speed.

While a number of early studies demonstrated that data accuracy deteriorates with response delay, the survey methodology literature offers few insights into how delay influences data completeness, response variance, or the presence of response set biases (Glasman and Ford 1988, Blair and Burton 1987, Cook 1987, Biderman and Lynch 1981). Differences in response quality are usually attributed to forgetting by respondents, a phenomenon that is most often assumed to increase with the amount of elapsed time between an event and its subsequent reporting.

Both experimental and non-experimental methodologists agree that delay poses risks of

maturation and history effects which can influence response variability, but the literature is ambivalent about the direction of these influences. Some researchers have argued that response variability increases with time (Hui and Triandis 1985, Hegenaars and Heinen 1982). According to these arguments, greater elapsed time between as event and its subsequent recall means greater opportunity for additional experience with the object of the event during the elapsed time and a greater likelihood of encountering more variability in product or service performance. Since recurring use of a product or service implies more-or-less continuous evaluation, the evaluation of the product under different circumstances over time leads to greater variability in performance evaluations on the different attributes.

On the other hand, there is reason to expect that response variability diminishes with response delay. Cognitive psychologists have proposed that information is stored in long-term memory according to its similarity to other information already held in storage. When information is recalled, the memory store is activated and missing or forgotten information is "filled in" with recalled information that is typical of past experiences (Ford and Smith 1987, Cohen 1981, Hamilton 1981). According to this view, the passage of time results in increased generalization across the attributes of a product or service category as well as increased generalization across multiple experiences with the service category (Alba and Marmorstein 1987, Blair and Burton 1987, Wyer and Srull 1981). Thus, when a person is asked to recall information about specific attributes of a complex product or service they purchased in the past, the elapsed time since the purchase affects the person's ability to discriminate. With longer elapsed times there would be less discrimination, greater similarity of scores across attributes and experiences, and a lower level of response variability (Burke and Srull, 1988). For a review of the cognitive demands associated with retrieval of information about past events see Jabine, et. al. (1984).

RESEARCH OBJECTIVES

The purpose of this study is to determine whether response delays affect the quality of satisfaction data in mail surveys. It is expected that response delays will adversely affect both data completeness and data usefulness. These expectations are formally stated as follows.

Hypothesis 1: Response delay will negatively affect satisfaction data completeness.
H1a: Late respondents will have a higher rate of section omission for satisfaction items than will early respondents.
H1b: Late respondents will have a higher level of item omissions than will early respondents.
Hypothesis 2: Response delay will negatively affect satisfaction data usefulness.
H2a: The level of response set bias will be higher for late respondents than for early respondents.
H2b: Satisfaction item response variance will be higher for late respondents than for early respondents.

RESEARCH METHODS

Setting and Respondents

Data used in this analysis were collected by means of a mail survey of new checking account customers of a large bank in the Southeastern United States. This survey was undertaken as part of an exploratory investigation of information search behaviors among new bank customers and their satisfaction with banking services.

Data Collection Procedure

The survey was administered in accordance with procedures advocated by Dillman (1978). A questionnaire was sent under a cover letter without pre-notification to 1005 persons who had opened checking accounts within two months of the initial mailing date. Postal Service records indicate that 948 of the 1005 questionnaires in the initial mailing were delivered. Three weeks after the initial mailing a second cover letter and replacement questionnaires were mailed to non-respondents. Within three weeks of the second mailing, 267 completed questionnaires had been returned, a response rate of 28 percent (Wiseman and Billington, 1984).

Comparison of information provided by

respondents to information available from bank records indicated no significant differences between respondents and non-respondents with regard to area of residence, account opening dates, or check writing frequency. Respondents did, however, make deposits to their checking accounts with a higher frequency than non-respondents. Although these differences were slight (2.39 deposits per month for respondents versus 2.09 deposits per month for non-respondents) they did prove to be statistically significant ($z = 2.40$; $p > z = 0.001$).

MEASURES

Items and Measures

The questionnaire consisted of three broad sections. The first contained behavioral items which asked respondents about search behaviors they had engaged in prior to choosing a bank. The second section contained scales on which subjects reported their satisfaction with bank services. The third section contained categorical questions asking for information about the respondent's age, income, and bank usage experience. A questionnaire completion date was also included among these questions.

Twenty-one bank service attributes were considered in the study. These attributes were selected after discussions with banking industry executives and consultants and included attributes from the SERVQUAL scale developed by Parasuraman, et. al. (1986). Respondents rated bank performance on each of the twenty-one attributes using seven-point scales ranging from "far below expectations" to "far above expectations".

Response Groups

Respondents were categorized as early or late according to a median split on questionnaire completion dates. The resulting groups were approximately equal with 133 respondents in the early group and 134 in the late group. Information available from bank records indicated that there were no statistically significant differences between the early and late groups in terms of area of residence, account opening dates, or check writing frequency.

RESULTS AND DISCUSSION

The early and late groups were compared in terms of section and item omissions, response set bias, and item variance. Section omission and response set bias were treated as dichotomous dependent variables.

Hypotheses involving these categorical dependent variables (Hypotheses 1a and 2a) were operationalized as tests of differences in proportions between the early and late response groups. Results of these tests are shown in Table 1.

Table 1
Tests of Differences in Proportions

	EARLY GROUP			LATE GROUP				
Variable	x	n	x/n	x	n	x/n	z-Value	p>z*
Section Omission	1	133	0.008	6	134	0.045	1.9051	0.0284
Response Set Bias	9	132	0.068	13	128	0.102	0.9669	0.1660

*under the one-tailed hypotheses that the observed level for the variable is higher for the late group than for the early group

Section Omission

A returned questionnaire was classified as having a section omission if no responses were provided for any of the 21 items in the satisfaction data section, and as not having a section omission if responses were provided for any of the 21 items. Only one of the 133 questionnaires (0.75 percent) returned by the early respondents contained section omissions for the 21-item satisfaction section. Six of the 134 questionnaires (4.48 percent) returned by the late respondents contained section omissions. Results of the statical tests for significance of observed differences provide support for the hypotheses. Table 1 shows that the z-value for the difference in observed proportions ($z = 1.9051$) is large enough to be significant at the 0.05 level ($p > z = 0.0284$) for the one-tailed test that the omission rate is higher for the late group than for the early group. Hypothesis 1a is,

therefore, supported.

Since questionnaires with section omissions, by definition, contained no responses to any of the 21 items, questionnaires with section omissions were excluded from subsequent analysis and hypothesis testing.

Response Set Bias

A returned questionnaire was classified as containing a response set bias if the same response was provided to all 21 items, and as not having a response set bias if there were any differences among responses to the 21 items.

Nine of the 132 questionnaires in the early response group (6.8 percent) contained a response set bias. This was lower than the proportion of late questionnaires containing response set bias (13 of 128, or 10.2 percent). The observed direction was as hypothesized, but the difference in response set bias was not statistically significant ($z = 0.9669$, $p > z = 0.1660$). Hypothesis 2a is, therefore, not supported.

Item Omission

Item omission was treated as a continuous dependent variable. In accordance with traditional methods, it was defined as the count of items, from the set of 21 items, for which no responses were provided (Durand, et. al. 1983). Results of the test of difference in mean number of item omissions is shown in Table 2.

Table 2
Tests of Differences in Means

Variable	Response Group	n	Group Mean*	z-value**	p>z***
Item Omission	Early	132	0.2481 (0.3083)		
	Late	128	1.2164 (0.3072)	4.17	0.0001

*Standard errors of group means appear to their right in parentheses.
**For testing the significance of observed differences in group means.
***Significance level for the observed value of z under the null hypothesis that mean scores are equal for the early and late groups.

Observed differences in item omission were in the hypothesized direction and were statistically significant. On average, the early respondents omitted about 0.25 of the 21 attribute items while the late respondents omitted an average of approximately 1.22 items. The z-value for the test of difference in mean number of omitted items was 4.17 which was significant at the 0.001 level. Hypothesis 1b is supported. The relatively low numbers of omitted items for both groups suggests that people who did not skip the section made an effort to provide responses to each item.

Item Variance

Comparisons were made of the variance of each of the twenty-one items by respondent groups. F-tests were used to determine the statistical significance of observed differences in response variance between groups. Results of these F-tests are presented in Table 3.

Results provide some support for the hypothesis that response variance is time dependent. Observed variance for nineteen of the twenty-one items was higher among late respondents than among early respondents. For the set of attributes as a whole, the average variance among the early respondents was lower than average variance among the late respondents (1.644 vs. 1.940; $t = 3.49$; $p > t = 0.01$). Increases in variance were large enough to be statistically significant at the 0.05 level for five of the nineteen items; "treats customers fairly", "capable employees", "sympathetic employees", "employees understand jobs", "and pleasant to deal with". Six additional attributes had increases in variance which were marginally significant ($p > F = 0.10$). Differences were not significant for either of the two attributes for which response variance was observed to decrease ("modern facilities", and "well dressed/neat employees").

Statistical significance was present for only those items measuring customers' satisfaction with the quality of their interactions with bank employees. This may be due to inherent differences in the variability of the underlying attributes. This seems plausible considering the fact that some attributes, such as bank facilities locations, hours of operation, or interest paid on

checking accounts, would not vary during the data collection period. Other attributes, particularly those reflecting the quality of services provided by bank employees, would be expected to vary during the data collection period due to differences between employees and differences within any given employee across time. The variability of service quality is recognized in the marketing literature, and it has been proposed that variability is an innate characteristic of services along with intangibility, perishability, and inseparability (Zeithaml 1981).

Table 3
Response Variance by Respondent Group for Each of Twenty-One Service Attributes

Item	Early Group Variance d.f.	Late Group Variance d.f.	F-Value
Individual Attention	1.8447 132	1.9827 127	1.1171
Automated Teller Machines	1.8739 130	2.0911 122	1.1890
Modern Facilities	1.3884 130	1.3215 125	0.9899
Treats Customers Fairly	1.3250 131	1.7920 125	1.4174**
Responsive Employees	1.7033 131	2.0907 127	1.2661*
Convenient Location	1.6181 132	2.0667 128	1.3172*
Capable Employees	1.4274 132	1.8435 126	1.3530**
Well Dressed/ Neat Employees	1.5080 131	1.4717 126	1.0147
Accurate Record Keeping	1.6712 132	2.0731 127	1.2893*
Sympathetic Employees	1.7735 132	2.2720 123	1.3748**
Pays Reasonable Interest	1.6660 132	1.8365 127	1.1457
Trustworthy Employees	1.3458 130	1.5992 127	1.2164
Respects Confidentiality	1.5557 131	1.5634 127	1.0366
Good Reputation/ Image	1.4927 132	1.8125 127	1.2620*
Keeps Time Promises	1.8415 132	2.3022 127	1.2994*
Employees Understand Jobs	1.4890 132	1.9666 127	1.3728**
Pleasant to Deal With	1.6461 132	2.1825 127	1.3780**
Up-to-Date Equipment	1.6510 131	1.7464 127	1.0903
Appropriate Facilities	1.4783 132	1.6797 126	1.1903
Convenient Hours	2.0520 132	2.2756 127	1.1526
Reasonable Fees and Charges	2.1593 131	2.7630 126	1.3304*

** $F(0.05; v1=v2=120) = 1.35$
* $F(0.10; v1=v2=120) = 1.17$

A second possible explanation for the differences in response variability is that respondents processed information about their interactions with bank employees differently from information about other bank attributes. Zeithaml (1981) used a typology first proposed by Nelson (1970) and later modified by Darby and Karni (1973) to describe service attributes. She suggested that consumers apply simple processes when evaluating search attributes for which information can be ascertained prior to purchase, but that they apply more complex processing when evaluating experience and credence attributes for which information can be discerned only after purchase or during consumption.

Information acquisition may also differ across attribute categories (Maute and Forrester 1989). It seems likely that people acquire enough information prior to opening an account to make the judgement that location, operating hours, and interest rates are satisfactory, then cease to search for information about these attributes. But it also seems likely that they would continue to acquire information to support ongoing evaluations of satisfaction with bank employees.

SUMMARY AND CONCLUSIONS

This study presents evidence that response delay affects both the completeness and variability of satisfaction data collected in mail surveys. Questionnaires returned by late respondents had significantly higher levels of section and item omissions than those returned by early respondents. These findings suggest that satisfaction researchers should pay at least as much attention to issues of data quality as they pay to response rate issues when collecting data in mail surveys.

Differences between early and late respondents' rates of response were not statistically significant. The fact that there were significant differences in other data quality measures indicates that researchers should use multiple indicators when assessing data quality and that reliance upon single indicators may lead to

conclusions which are simplistic and inaccurate.

Responses of late respondents showed significantly higher levels of item variance than those of early respondents for some items, but not for others. Differences in item variance were present for only those attributes which reflected respondents' satisfaction with their interactions with bank employees. While this paper has offered some plausible explanations for this phenomenon, further research is needed to determine how the effects of response delay on response variance are moderated by attribute qualities. Results of such research will enable satisfaction researchers to be more effective in guarding against the potentially damaging effects of data degradation which accompanies response delay.

REFERENCES

Alba, Joseph W. and Howard Marmorstein (1987), "The Effects of Frequency Knowledge on Consumer Decision Making," *Journal of Consumer Research*, 14(Jun.), 14-25.

Biderman, Albert D. and James P. Lynch (1981), "Recency Bias in Data in Self-Reported Victimization." *Proceedings of the American Statistical Association.*

Blair, Edward and Scot Burton (1987), "Cognitive Processes Used by Survey Respondents to Answer Behavioral Frequency Questions," *Journal of Consumer Research*, 14(Sep.), 280-288.

Burke, Raymond R. and Thomas K. Srull (1988), "Competitive Interference and Consumer Memory for Advertising," *Journal of Consumer Research*, 15(Jun.), 55-68.

Cohen, Claudia E. (1981), "Goals and Schemata in Person Perception: Making Sense from the Stream of Behavior," in N. Cantor and J.F. Kilstrom (eds.), *Personality, Cognition, and Social Interaction*, Hillsdale, NJ: L. Erlbaum Associates, Inc.

Cook, William A. (1987), "Telescoping and Other Memory Tricks," *Journal of Advertising Research*, 27 (Feb./Mar.) RC5-RC8.

Darby, M. R. and E. Karni (1973), "Free Competition and the Optimal Amount of Fraud," *Journal of Law and Economics*. 16(April): 67-86.

Dillman, Don A. (1978), *Mail and Telephone Surveys: The Total Design Method*, New York: NY, John Wiley and Sons, Inc.

Dommeyer, C.J. (1989), "Offering Mail Survey Results with a Lift," *Journal of the Market Research Society*, 31 (3) 399-408.

Dommeyer, C.J. (1988), "How Form of the Monetary Incentive Affects Mail Survey Response," *Journal of the Market Research Society*, 30 (3) 379-385.

Duncan, Otis D. and M. Stenbeck (1989), "No Opinion or Not Sure?," *Public Opinion Quarterly*, 52 (Winter) 513-525.

Durand, Richard M. and Zarrel V. Lambert (1988), "Don't Know Responses in Mail Surveys: Analyses and Interpretational Consequences," *Journal of Business Research*, 16 (Mar.) 169-188.

Durand, Richard M., H. J Guffey, Jr. and J. M. Planchon (1983), "An Examination of the Random versus Nonrandom Nature of Item Omissions," *Journal of Marketing Research*, 20 (Aug.) 305-313.

Ford, Gary T. and Ruth Ann Smith (1987), "Inferential Beliefs in Consumer Evaluations: An Assessment of Alternative Processing Strategies," *Journal of Consumer Research*, 14 (Dec.), 363-371.

Glasman, Myron and John B. Ford (1988), "An Empirical Investigation of Bogus Recall," *Journal of the Academy of Marketing Science*, 16 (Fall) 38-41.

Goetz, Edward G., Tom R. Tyler, and Fay L. Cook (1984), "Promised Incentives in Media Research: A Look at Data Quality, Sample Representativeness, and Response Rate", *Journal of Marketing Research*. 21 (May) 148-154.

Goldsmith, Ronald E. (1989), "Reducing Spurious Response in a Field Survey," *Journal of Social Psychology*, 129 (Apr.), 201-212.

Goldsmith, Ronald E. (1988), "Spurious Response Error in a New Product Survey," *Journal of Business Research*, 17 (Nov.) 271-281.

Goudy, Willis J. (1976), "Nonresponse Effects on Relationships Between Variables," *Public Opinion Quarterly*, 40 (3) 360-369.

Groves, Robert M. (1987), "Research on Survey Data Quality," *Public Opinion Quarterly*, 51 (Winter) Part 2, S156-S172.

Hamilton, David L. (1981), "Cognitive Representations of Persons," in E.T. Higgins, C.P. Herman, and M.P. Zanna (eds.), *Social Cognition: The Ontario Symposium*, Hillsdale, NJ: L. Erlbaum Associates, Inc.

Hawkins, Del I., Kenneth A. Coney, and Donald W. Jackson, Jr. (1988), "The Impact of Monetary Inducement on Uninformed Response Error," *Journal of the Academy of Marketing Science*, 16 (Summer) 30-35.

Hegenaars, Jacques T. and Ton C. Heinen (1982), "The Effects of Role Independent Interviewer Characteristics on Responses", in Djikstra, W. and J. van der Souwen (eds.) *Response Behavior in the Survey Interview*, London: Academic Press.

Hui, C. Harry and Harry C. Triandis (1985), "The Instability of Response Sets", *Public Opinion Quarterly*, 49: 253-260.

Jabine, Thomas B., M. L. Straf, J.M. Tanur, R. Tourangeau (1984), *Cognitive Aspects of Survey*

Methodology: Building a Bridge Between Disciplines. Washington, DC: National Academy Press.

Maute, M. F. and W. R. Forrester, (1989), "A Preliminary Investigation of Information Search Processes for Complex Services", *Proceedings of the XIVth Annual Colloquium of the International Association for Research in Economic Psychology*, T. Tyszka and P. Gasparski (eds.), pp. 159-175.

Nelson, Philip (1970), "Information and Consumer Behavior," *Journal of Political Economy*, 78 (20), 311-329.

Parasuraman, A., Valarie A. Zeithaml and Leonard L. Berry (1986), "SERVQUAL: A Multiple-Item Scale for Measuring Customer Perceptions of Service Quality", Boston, MA: Marketing Science Institute. Report No. 86-108.

Rubens, William (1989), "We Don't Care About Research Quality Anymore," *Journal of Advertising Research*, 29 (Feb./Mar.), RC3-RC6.

Wiseman, Frederick, and Maryann Billington (1984), "Comment on a Standard Definition of Response Rates", *Journal of Marketing Research*, 21 (August) 336-338.

Wunder, Gene C., and George W. Wynn (1988), "The Effects of Address Personalization on Mailed Questionnaires Response Rate, Time, and Quality," *Journal of the Market Research Society*, 30 (1) 95-101.

Wyer, Robert S., and Thomas K. Srull (1981), "Category Accessibility: Some Theoretical and Empirical Issues Concerning the Processing of Social Stimulus Information," in E.T. Higgins, C.P. Herman, and M.P. Zanna (eds.), *Social Cognition: The Ontario Symposium*, Hillsdale, NJ: L. Erlbaum Associates, Inc.

Zeithaml, Valarie A. (1981), "How Consumer Evaluation Processes Differ between Goods and Services." In *Marketing of Services*, James H. Donnelly and William R. George, eds., Chicago, IL: American Marketing Association: 191-199.

Send correspondence regarding this article to:
William R. Forrester, Jr.
Department of Marketing
Kennesaw State College
Marietta, GA 30066