Probability of Selecting Response Mode Between Paper- and Web-Based Options: Logistic Model Based on Institutional Demographics

Marina Murray

Graduate Management Admission Council , 11921 Freedom Drive, Suite 300, Reston, VA, 20190

Abstract

During the last decade, a rapid growth in online surveys supported by advances in survey software technology and increased Internet usage inspired a great deal of methodological research to gain a better understanding of web survey mode. Most of this research, however, focuses on individual respondents who participate in surveys to the general public. Are online surveys ready to replace paper-and-pencil formats for the purpose of studying organizations? When organizations have a choice, which method do they prefer? What factors influence a company's decision in the survey mode selection? The findings addressing these questions are based on primary research data collected by the Graduate Management Admission Council through its 2006 Diversity Survey of business schools in the United States. In addition to the traditional paper-and-pencil survey format, these US business schools were offered the alternative of completing the survey online. Using logistic regression, this research aims to estimate which institutional demographic characteristics influence the survey mode choice in establishment surveys. Depending on institutional demographics, which are typically known to the researcher, the results may be helpful in choosing which survey mode to utilize when conducting survey research of organizations.

Key Words: multimode surveys, establishment surveys, logistic regression

1. Introduction

The Internet provided survey researchers with a powerful array of tools and means to lower the cost of fielding surveys, to shorten data collection time (Couper, 2000), to have immediate access to data (Ilieva, Baron, & Healey, 2002), and to reduce an overall survey error (Dillman, 2009), among others. For example, with nearly all organizations in the United States having Internet access via high-speed lines (FCC, 2009), the likelihood of under-coverage error attributed to the online survey mode in establishment surveys appears to be much lower than that in general population surveys. Yet, whether driven by tradition, record-keeping convenience, sampling considerations, or respondent preference, traditional mail surveys remain one of the preferred methods of surveying organizations in the United States (Dillman, 2007). Nonetheless, technological developments and increased demand for faster turn-around times and lower costs (Macer, 2003) are moving establishment surveys towards multi-mode (de Leeuw, 2005) or web-only data collection methods.

The purpose of this paper is to identify factors influencing the response mode choice in establishment surveys and contribute to better understanding of mixed-mode issues in survey methodology.

2. Background

2.1 Response Rates in Mixed-Mode Surveys
Offering choice between modes sequentially typically increases response rate (Dillman, Tortora, Swift, Kohrell, Berck, & Messer, 2009). When provided with a simultaneous choice of modes, however, the response rate can be the same (Porter & Whitcomb, 2007) or lower (Israel, 2010) compared with one-mode design. On the other hand, choice may increase respondent motivation to complete a survey (Bäckström & Nilsson, 2002). Kaplowitz, Hadlock, & Levine (2004) explored the effect of the survey mode on response rates and found that beginning the survey with a paper letter and following up with a web follow-up increases response rates.

2.2 Binary Logistic Regression Model
There is a vast spectrum of research that illustrates logistic regression model application in predicting a binary response based on a set of independent variables in social science research (Hosmer & Lemeshow 2000; Menard, 2002, Archer & Lemeshow 2006). With fewer assumptions to meet compared to other regression models, e.g., OLS, and a wider range of acceptable predictors (interval, ratio, dichotomous), the model works well with survey data.

3. Methodology

Analysis presented in this paper is based on primary research data collected by the Graduate Management Admission Council (GMAC) through its 2006 Diversity Survey. This, was a survey of graduate business schools in the United States using GMAT exam scores that expressed interest in information related to diversity initiatives. The survey purpose was to learn how business schools addressed racial diversity in their organizations, what efforts they undertook, what challenges they faced, and what best practices were emerging. This survey was originally designed as web-only and launched in June 2006. Even with a follow-up email, only 31 of 647 schools responded by the one-month cut off date, for a 4.8% response rate.

The study was redesigned to give invitees the option of responding via a paper-and-pencil or on-line. A new mailing to the non-respondents was sent in October 2006. , respondents had a choice of completing a paper survey and sending it back in the enclosed business reply envelope or submitting data via web using an easy URL and individual respondent key that was included in each questionnaire. Two weeks were allowed for mailing and one month for response. By the new response cut-off date (tracked by a postmarked stamp for paper questionnaires), 187 schools responded at 28.9% response rate. This response was achieved after one follow-up mailing and represents a significant improvement (24 percentage points) compared with the initial web-only design. Of participating institutions, 110, or 60.1%, responded online. Data submitted by multiple programs within one institution were aggregated on the institutional level.

3.1 Survey Instrument
The survey instrument consisted of 23 detailed questions. Based on the pilot test results, the average paper survey completion time was 10 minutes. For those who responded to

the online survey, it took on average 7 minutes. Broad survey topics included diversity initiatives, minority-oriented financial aid programs and student services, and affirmative action legislation. Care was taken to ensure a similar look and feel between paper and web versions of the questionnaire. However, skip patterns were programmed into the online survey and compliance was enforced for selected questions.

3.2 Demographic Variables

The following institution demographic characteristics were collected as part of the Diversity Survey: institution governance status (public or private), location (state), type of urbanization (city, urban fringe, town, or rural area), the number and type (full-time, part-time, executive, doctoral) of programs offered, program management (centralized or decentralized). The following variables were retrieved from other GMAC data and merged with the respondent data file: average class size, program-based annual growth rate, and number of GMAT scores received in testing year 2006. The following publicly available data were also merged with the data file: university size (the number of students applied, the number of students admitted, the number of students enrolled), college affordability index, published full-time MBA program rankings (BusinessWeek, Financial Times, US News & World Report). Selected variables were studied as is while values of other variables were as well grouped into broader categories to either address a working hypothesis or for the predictor scale range comparability.

4. Results

The first analysis examined each of the variables independently to determine whether each variable accounted for significant differences in the response rates. The results are shown in Table 1. The variables are grouped in Table 1 to indicate which variables were later included in the logistic regression model.

Contrary to the initial assumption that schools in the US West would be more likely to respond online, the opposite trend was observed. A large and significant difference between schools responding online versus completing mail surveys was observed in the average number of GMAT score reports received.

Variables describing institutional demographic characteristics at the university level rather than the business school or at program levels were excluded from the logistic regression. The number of GMAT scores received was also excluded to avoid multicollinearity between predictors (the number of score reports received was highly correlated with program rankings). One advantage of dropping that variable is that the resultant set of variables for model inclusion was all available either through the survey or through publicly available sources, and not through proprietary databases.

Table 1: Univariate Analysis of Selected Variables Studied

Variable	Web	Paper	p
Not in the model			
West (indicator)	13.6%	26.0%	**< .05**
Large city (indicator)	35.5%	42.5%	> .05
Number of GMAT score reports received	912.6	226.9	**< .05**
Number of graduate business programs offered	4.8	4.8	> .05
Centralized program management (indicator)	51.6%	60.7%	> .05
Program growth rate	6.8%	13.8%	**< .05**
Public (indicator)	48.2%	60.3%	> .05
University size (enrolled)	2,373	2,006	> .05
In the model			
Ranking group (3rd-tier =1 to 1st tier=3, first tier presented)	13.6%	5.5%	**< .05**
Class size (indicator: (100 or more students=1)	64.5%	53.4%	> .05
Program delivery type (indicator: various=1, full-time only=0)	63.6%	46.6%	**< .05**
Fields of study (various=1, MBA only=0)	73.6%	46.6%	**< .05**
Graduate degree types (various=1, master-level only=0)	94.5%	82.2%	**< .05**

A logistic regression analysis based on iterative maximum likelihood algorithm was employed using paper response versus web response as the dependent variable. Table 2 displays the results from this analysis. There was one ordinary variable (ranking group[1]) and four dichotomous variables (class size, program delivery type, fields of study, degree types) among predictors. A test of the full model versus a model with only one intercept was statistically significant (χ^2= 45.66, df = 5, p = 0.00).

Table 2: Logistic Regression Predicting Response Mode Selection Between Paper and Web Surveys

Variable	β	Wald	Sig	Exp (β)	95%CI upper	95%CI lower
Ranking group	.87	7.88	.005	2.39	1.30	4.38
Class size	.82	5.21	.022	2.28	1.12	4.62
Program delivery types	.93	6.29	.012	2.53	1.22	5.22
Fields of study	1.55	17.36	.000	4.69	2.27	9.71
Degree types	2.27	14.57	.000	9.65	3.01	30.89
Constant	-4.68	25.63	.000	0.01		

Correct predictions: 76.1%
Omnibus test of model coefficients: χ^2 = 45.66, df = 5, p = 0.00
Model summary:
    -2 Log Likelihood = 200.50
    Cox & Snell R^2 = 0.22
    Nagelkerke R^2 = 0.29
Hosmer and Lemeshow goodness-of-fit: χ^2 = 1.36, df = 8, p = 0.99

---

[1] Computed based on the average media rankings of full-time MBA programs across *BusinessWeek, Financial Times*, and *U.S. News & World Report* where the 50 top-ranked institutions were grouped into the first tier (coded "3"), schools ranked 50 to 100 were grouped into the second tier ("2"), and those ranked 101 or higher as well as nonranked schools were in the third tier ("1").

The model was able to classify correctly 81.8% of those who responded online and 66.2% of paper respondents for an overall percentage of 76.1% for correct predictions. A nonsignificant χ^2 in the Hosmer-Lemeshow test indicates that the data fit the model well.

Four of the five variables included in the model yielded significant differences between the web and paper groups. The web group consisted of business schools that were more likely to have higher published rankings and offer a variety of degree types and program delivery types in multiple fields of study.

Wald χ^2 for each individual predictor is statistically significant and confirms their individual contribution to the model. The odds ratio (Exp (β)) indicates that given all other predictors constant, institutions offering a variety of degree types were nearly 10 times more likely to respond online than those offering less variety of degree types. Schools that offered a variety of fields of study were nearly 5 times more likely to respond online than those that did not. One point of increase in the ranking group, e.g., from the second tier to the first tier, more than doubles the probability of the web response.

5. Discussion

In this study, when given a simultaneous choice, schools with a broader scope of program offerings were more likely to respond to a survey via the Internet. On the other hand, program size revealed an opposite trend—larger programs tended to respond via paper. It appears that the level of an organization's diversification may be a strong indicator of its preferred response mode. Introducing an additional response mode generated a better response rate than using a web-only or paper-only design, which confirmed similar research findings (e.g. Dillman, et.al., 2009).

The study has several limitations that can be addressed in future research. First, it was limited to a narrow population of graduate business schools in the United States that both received GMAT score reports and expressed interest in diversity-related topics. Second, the sensitivity of survey topics not measured in this study may have affected respondent choice. Third, institution demographics were limited to survey data and publicly available data and may not have included characteristics that strongly impact the mode response choice. Fourth, the initial email invitation to a web-only survey may have been considered as a prenotification, which is known to increase response rates; thus turning the empirically based adjustment into a sequential study design. Last, the variety of offerings or diversification is likely to be a function of a latent characteristic not captured by institution size and hence not examined in this study.

References

Archer, K.J., Lemeshow S. (2006). Goodness-of-fit test for a logistic regression model fitted using survey sample data. *The Stata Journal*, *6*(1), pp. 97–105.

Bäckström, C and Nilsson, C (2002) Mixed mode: handling method differences between paper and web questionnaires. Retrieved from http://www.modsurvey.org/text/MixedMode-MethodDiff.pdf.

Couper, M.P. 2000. Web surveys: a review of issues and approaches. *Public Opinion Quarterly*, *64*, pp. 464–494.

de Leeuw, E.D. 2005. To mix or not to mix: data collection modes in surveys. *Journal of Official Statistics, 21*(2), pp. 233–255.

Dilman, D.A. (2007). Mail and Internet surveys the tailored design method (2nd ed.). Hoboken, NJ: Wiley & Sons.

Dillman, D.A., Phelps, G. Tortora, R., Swift, K. Kohrell, J. Berck, J., & Messer, B. Response rate and measurement differences in mixed mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. Retrieved from http://www.sesrc.wsu.edu/dillman/papers.htm.

Dillman, D.A., Smyth, J.D., & Christian, L.M. 2009. *Internet, mail, and mixed‑mode surveys: the tailored design method* (3rd ed.). Hoboken, NJ: Wiley & Sons.

Federal Communications Commission (FCC) (2009). *A national broadband plan for our future.* Washington, DC: FCC.

Kaplowitz, M.D., Hadlock, T.D., & Levine, R. 2004. A comparison of web and mail survey response rates. *Public Opinion Quarterly, 68*(1), pp. 94–101.

Hosmer, D.W., Lemeshow, S. 2000. *Applied logistic regression* (2nd ed.). Hoboken, NJ: Wiley & Sons.

Israel, G.D. 2010. Using web-hosted surveys to obtain responses from extension clients: a cautionary tale. *Journal of Extension, 48*(4).

Ilieva, J., Baron, S., & Healey, N.M. 2002. Online surveys in marketing research: pros and cons. *International Journal of Market Research, 44*(3), pp. 361.

Macer, T. 2003. We seek them here, we seek them there. How technical innovation in mixed mode survey software is responding to the challenge of finding elusive respondents. Retrieved from http://www.meaning.uk.com/your-resources/papers-and-presentations.

Menard, S.W. (2002). Applied logistic regression analysis. Thousand Oaks, CA: Sage Publications.

Porter, S.R., & Whitcomb, M.E. 2007. Mixed-mode contacts in web surveys: paper is not necessarily better. *Public Opinion Quarterly*, 71(4), p. 635.