Increasing both survey completion rates and data quality remains an important topic for fields as diverse as sociology, marketing, medicine and history. Thousands of studies have made response quality their central topic of examination, but their focus has largely been to measure response bias through the comparison of early–late wave responses. In this study, an innovative online field experiment tests a two-staged highly interesting question to produce an 8% better survey completion rate and to change sample representativeness by 12% over a usual one-stage highly interesting question appearing at the beginning of the questionnaire. In addition to these substantive findings, a distributional and probability analysis is developed that further refines methods for identifying the extent of non-response bias.

Introduction

The relationship between respondent interest in a survey topic and the survey response was hypothesised more than half a century ago (Blankenship et al. 1949, p. 410) and is at times referred to as the 'interest hypothesis' (e.g. Armstrong & Overton 1977; Wright & Armstrong 2008). Conversely, refusal to participate in surveys often results from low topic interest (e.g. Groves et al. 2000). It is therefore not surprising that survey topic and content have been documented as typically the most important factors in stimulating response rate (Greer et al. 2000).
Survey response rates have been falling for the last several decades (see, e.g., Bickart & Schmittlein 1999; De Leeuw & De Heer 2002; Jarvis 2002), giving rise to concerns about decreased research generalisability, bias in sample estimates, and information loss due to increasingly longer surveys (e.g. Adiguzel & Wedel 2008). And a recent meta-analysis indicates the response rate is normally 11% lower for web surveys versus mail surveys (Manfreda et al. 2008). While the offering of monetary rewards or other incentives has been the typical approach to boosting response rates for lengthy or less interesting research studies (e.g. Groves et al. 2000; Groves et al. 2004), recent research indicates that even incentives may not be enough (Childers & Skinner 1996; Gofton 1999; Sheehan 2001; Yankelovich Monitor 2004; Hulme 2005; Porter 2005). As such, increasing response rates and decreasing potential non-response error are important issues in survey research (Collier & Bienstock 2007, p. 179).

The objective of this article is to examine how a two-stage questioning approach combined with a highly interesting question affects response rates and potential non-response bias. Additionally, methodological contributions are made by describing and applying distributional and probability analysis to further refine methods for identifying the extent of non-response bias.

'Highly interesting questions' are generally hot media topics that (when used as part of a survey) are intended to increase participant interest by increasing value in the reciprocity of exchange; sometimes they are also called 'inducement questions' (e.g. Geurts et al. 1988; Geurts & Whitlark 1994). The questions may be about emotionally charged topics related to public policy, court cases, celebrity happenings, or national or geographic events, etc. In all cases, it is important that respondents view the highly interesting question as an integrated part of the survey whole to communicate the legitimacy of the survey.

The current study compares one online survey that begins with a highly interesting question, with a second otherwise identical online survey containing a two-stage version of the same highly interesting question (Figure 1). The two-stage question topic is (1) mentioned at the beginning of the survey, but is (2) answered towards the end of the survey (i.e. a type of foot-in-the-door approach) (Figure 2).

The study thereby measures the difference in participation completion rate when answering the different highly interesting questions (Hypothesis 1), and the extent to which non-participation (non-response bias) affects the generalisability of survey results (Hypotheses 2, 3a and 3b).
In regards to funding future social security, do you feel your taxes should be:
- Not changed
- Raised
- Lowered

At the end of the questionnaire, there will be an additional question regarding how much you feel your taxes should be raised or lowered.

Then, at the end of the survey ...

In regards to funding future social security, if you answered that your taxes should be raised or lowered, then by what per cent?
- 2.5%
- 5%
- 7.5%
- 10%

**Figure 1** Example two-stage highly interesting question

<table>
<thead>
<tr>
<th>One-stage approach</th>
<th>Questionnaire completion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answer highly interesting question (HIQ)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proposed two-stage approach</th>
<th>Questionnaire completion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mention HIQ topic</td>
<td></td>
</tr>
<tr>
<td>Complete the main body of the survey</td>
<td></td>
</tr>
<tr>
<td>Answer highly interesting question (HIQ)</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2** Two-stage approach to highly interesting questions

**Background**

Survey participation refusal often results from low topic interest (e.g. Dillman 1991; Martin 1994; Groves *et al.* 2004). Leverage-salience theory suggests survey participation increases with the personal interest in the survey. Groves *et al.* (2004, p. 25) find ‘people cooperated at a higher rate to surveys on topics of likely interest to them’. In short, people are more likely to answer questions they find interesting. Common survey research wisdom is consistent with the conclusion made by Blankenship and co-authors (1949, p. 410) over half a century ago that ‘those who reply are probably not typical members’ of the intended survey population but rather are ‘those especially interested, or those particularly in opposition to the ideas presented’. Also well established are the effects of question
Two-stage highly interesting questions in online marketing research

location on the response rate to individual questions. The survey literature abounds with examples demonstrating that responses are sensitive to the placement of the questions. DeMoranville and Bienstock (2003) provide a more complete review of literature related to these question order effects.

To date, however, researchers have not considered the location effects of a highly interesting question on the response rate or distributional quality of total survey responses. We propose that the two-stage highly interesting question produces a set of completed responses that are less extreme (i.e. less interested or less opposed to the survey as a whole), because, compared to the group completing the one-stage (same) highly interesting question survey, some are less interested in the overall survey (e.g. have more moderate views on it) and complete it only to get to the second stage of the highly interesting question. Thus:

**H1:** A two-stage highly interesting question in a survey results in a higher completion rate at survey conclusion than a corresponding one-stage highly interesting question.

**Identifying non-response bias**

There is a longstanding concern about survey non-response bias that is scattered across journals in a variety of fields, including marketing, health, political science, psychology, economics and applied statistics (Singer 2006; Miller 2008, p. 4). One commonality in opinion research across these fields is the convention of assessing non-response bias through comparing mean scores of early and late respondents. Indeed, a review of articles finds at least 1500 survey-based research articles conducting such a test. According to Wright and Armstrong (2008), much of this focus is sourced to widespread misinterpretation of a 1977 article by Armstrong and Overton on survey non-response. Likewise, Collier and Bienstock (2007, p. 180) suggest that researchers must ‘question how valid it is to extrapolate early and late responders in an online setting’. The online data collection task with panels is often reduced to less than 24 hours and offers little time differentiation. However, no published empirical research has reported the extent to which conventional statistical testing of early vs late mean scores is actually able to identify the presence of non-response bias.

As an alternative approach to identifying and estimating non-response bias, we propose that the one-stage and two-stage surveys' mean scores are not different, but that their answer distributions are different. If we consider an early wave of respondents who are largely ‘especially interested’
(e.g. 7s) or ‘particularly opposed’ (e.g. 1s) on a seven-point scale, and a later wave of respondents are ‘less interested’ (e.g. 5s) or ‘less opposed’ (e.g. 3s), we should observe roughly the same mean score (e.g. a 4). Thus:

**H2:** There is not a statistically significant difference in variable mean scores between concurrent surveys using one-stage versus two-stage highly interesting questions.

However in contrast to the ‘especially interested’ or ‘particularly opposed’ respondents that typically reply to surveys, respondents who would otherwise not participate – but who take the survey to arrive at the late-appearing highly interesting question – should be less interested or less opposed as to the overall survey. This latter group should be characterised by a more normal multivariate answer distribution with an increased number of centre-scale answers (rather than a more extreme platykurtic distribution expected for respondents to the one-stage highly interesting question) (see Figure 3).

Consistent with this logic, we propose that the two-stage highly interesting question survey contains a more normal distribution (on both ends of the scale) – as indicated by a smaller standard deviation score and associated smaller platykurtic score – representing additional respondents who otherwise would not participate in the survey, and who reduce non-response bias. Thus, whereas hypothesis 2 tests for differences in mean scores, hypothesis 3 tests for differences in variance. Stated formally, it takes the form:
H3a: The distributions of respondents’ answers in a survey with a two-stage highly interesting question have a smaller standard deviation than the distributions of respondents’ answers in a concurrent survey with a corresponding one-stage highly interesting question.

H3b: The distributions of respondents’ answers in a survey with a two-stage highly interesting question are less platykurtic than the distributions of respondents’ answers in a concurrent survey with a corresponding one-stage highly interesting question.

Methods

Central to this article’s focus is the tracking and analysing of survey completion rates on a question-by-question basis that online surveying permits (Bosnjak & Tuten 2001; Bosnjak et al. 2005). The experimental design of this paper uses concurrent waves suggested by Armstrong and Overton (1977) – comparing responses to surveys with one-stage and two-stage highly interesting questions – to provide an important, first investigation of the extent to which different statistical tests are capable of detecting potential non-response bias.

The final sample of \( n = 712 \) was obtained from a commercial online database broker through the online survey software provider, Qualtrics.com. Approximately 4000 respondents were randomly selected from a double opt-in national panel of individuals age 18 or older with a unique email address. The sample was drawn to reflect known proportions of region, age and gender in the United States following probability-based sampling techniques (Couper 2000).

To ensure initial respondent uniformity between the two concurrent surveys, all 4000 potential respondents were sent the same invitation via their unique email addresses. This identical invitation, following normal survey language, also described the topics of the survey (i.e. Wal-Mart, Ford, President Bush) and the highly interesting question (i.e. on President Bush raising taxes to fund social security). When participants clicked on the hyperlink contained in the email message, a web-based JavaScript randomly assigned and redirected each participant to one of the two surveys: either the one-stage or two-stage highly interesting question survey. The JavaScript was invisible to participants, and transfer was virtually instantaneous. One week after sending out the initial invitation, a reminder invitation was sent out to potential respondents who had not previously clicked on the
email hyperlink. The reminder invitation contained the same information and hyperlink, and second-wave respondents continued to be randomly assigned to the two survey conditions.

The final count revealed samples of 347 respondents who initially visited the one-stage highly interesting question survey and 365 respondents who initially visited the two-stage highly interesting question survey, for a total sample of 712 respondents between the two surveys (out of the 3992 individuals – or 18% of those – who were sent the identical survey email invitation). The respondent demographic profiles included age, ethnicity, education and household income measures, and were consistent with recent research on survey response (e.g. Groves et al. 2004; De Rada 2005). Each was collected as part of the survey and measured using standard categorical responses. We note the respondents were in proportion to population across most of the United States.

The survey was developed and refined through extensive pretesting of students, faculty and professionals. The survey consisted of 85 questions containing 5-point Likert, semantic differential and anchored sum questions – spread across six survey web pages. All responses were validated such that the respondent could not proceed to the next page until all questions were answered.

The survey questions were developed around four measures of product/brand leadership used extensively in prior marketing and other social science literature (Cattell & Drevdahl 1955; Digman 1990; Kirkpatrick & Locke 1991; Saucier 1994; Aaker 1997; Brown et al. 2002). These measures included four leadership measures that asked about the subjects being: Good for America, Innovative, Likeable and Competent. As the subject topics for these leadership measures, Wal-Mart Stores, Ford Motor Company and President George W. Bush were chosen as each was being discussed by the news media at the time of the survey, and respectively represented consumer products, durable products, and services. The subjects were evaluated on the leadership measures using a 5-point rating scale having descriptive tails with 1 = does not describe at all, 5 = describes perfectly well. These 12 measures (4 × 3) provided the variables for analysing completion rates, mean scores and kurtosis levels, and were selected to provide a meaningful examination of variance in responses (i.e. that reflects application differences in retail, manufacturing and executive leadership). The media was also discussing potential legislation by George W. Bush on reform of the US social security system. Given the general public interest in the topic, the last topic (related to a portion of the survey on service leadership traits) was chosen as the highly interesting question topic.
The complex question of 'Who would have refused to answer the survey had they been in the one-stage HIQ instead of the two-stage HIQ' requires a complex answer/analysis, given that 'it is usually impossible to compare respondents and nonrespondents on precisely those variables of most survey interest' (Dillman 1991, p. 229). To answer this question, the analysis must not be biased and exclude the incremental response group from the initial test(s). This is a complex analysis, given that researchers do not know at the beginning which individuals from the response group would have responded if they were in the control group. To predict which respondents in the treatment survey represent the difference in refusal rate at completion between the two surveys, a subgroup probability analysis is conducted using a combination of iterative clustering, discriminant analysis, general linear modelling, matrix multiplication and chi square distribution analysis (Figure 4).

First, iterative clustering is used to segment all respondents in the 'control group' (i.e. the survey with one-stage highly interesting question) based on their answers to the 12 questions on subject-trait scores (e.g. Wal-Mart—good for America to President Bush—competent) into natural response patterns. Next, discriminant analysis is used on the control condition with the clusters as the dependent variables, and the 12 descriptive survey items as the independent variables to create probability indices. Then, general linear modelling is performed using the six probabilities from the control condition to calculate beta coefficients for the 12 questions based on the control group respondent scores. The beta coefficient matrix from the control group is then multiplied by the individual respondent scores from the treatment group (i.e. the survey with the two-stage highly interesting question), resulting in cluster probability levels for each respondent in the treatment group. The procedure was repeated using four, five, seven and eight clusters for robustness.

Figure 4 Subgroup probability analysis overview
Findings

Comparability of the survey respondents

As to demographics, the two surveys are comparable overall across age, income and education distributions (see Table 1). However the incremental respondents (identified through the subgroup probability analysis in a later section) are younger, less educated and have less income. Additionally, respondents spent near equal time-to-completion between the two

Table 1  Highly-interesting question location and variable demographic distribution percentages

<table>
<thead>
<tr>
<th>Variables</th>
<th>One-stage vs two-stage</th>
<th>... vs subgroup</th>
<th>Sig1</th>
<th>Sig2</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (completed)</td>
<td>173</td>
<td>211</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time to completion (in minutes)</td>
<td></td>
<td></td>
<td>0.94</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>1st quartile</td>
<td>7.3</td>
<td>7.8</td>
<td>7.4</td>
<td>7.4</td>
</tr>
<tr>
<td>2nd quartile</td>
<td>9.8</td>
<td>10.0</td>
<td>9.8</td>
<td>9.8</td>
</tr>
<tr>
<td>3rd quartile</td>
<td>13.0</td>
<td>13.3</td>
<td>13.0</td>
<td>13.0</td>
</tr>
<tr>
<td>Mean</td>
<td>11.8</td>
<td>11.9</td>
<td>0.97</td>
<td>11.7</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) 18–24 years (%)</td>
<td>8</td>
<td>10</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>(2) 25–34 years (%)</td>
<td>18</td>
<td>18</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>(3) 35–44 years (%)</td>
<td>26</td>
<td>27</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>(4) 45–54 years (%)</td>
<td>30</td>
<td>27</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>(5) 55–64 years (%)</td>
<td>13</td>
<td>13</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>(6) 65+ years (%)</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td>0.95</td>
<td>0.01</td>
</tr>
<tr>
<td>(1) Some high school (%)</td>
<td>6</td>
<td>6</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>(2) High school graduate (%)</td>
<td>22</td>
<td>21</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>(3) Some college (%)</td>
<td>38</td>
<td>39</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>(4) College graduate (%)</td>
<td>22</td>
<td>26</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>(5) Postgraduate/professional (%)</td>
<td>13</td>
<td>9</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Income</td>
<td></td>
<td></td>
<td>0.83</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>(1) $14,999 or less (%)</td>
<td>16</td>
<td>16</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>(2) $15,000–$24,999 (%)</td>
<td>14</td>
<td>18</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>(3) $25,000–$39,999 (%)</td>
<td>20</td>
<td>22</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>(4) $40,000–$54,999 (%)</td>
<td>19</td>
<td>18</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>(5) $55,000–$79,999 (%)</td>
<td>18</td>
<td>17</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>(6) $80,000 or more (%)</td>
<td>13</td>
<td>10</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Notes:
1 Demographic summary statistics are comparable across the two surveys. 2 The subgroup scores, in contrast, show significant differences in the demographic between the one-stage survey respondents and the subgroup of respondents in the two-stage survey (identified separately through probability analysis of descriptive statistics as described in the text). Taken together, the findings indicate that demographic variables are not capable of always identifying non-response bias in descriptive data.
Two-stage highly interesting questions in online marketing research

experimental conditions ($p < 0.971$). These results are consistent with other research confirming the difficulty of predicting who completes a survey based on aggregate demographics alone.

**Completion and refusal rates**

As shown in Figure 5, the test of the two-stage highly interesting question survey had an 8% higher completion rate when compared to the otherwise identical one-stage highly interesting question survey, thus supporting Hypothesis 1 ($p < 0.033$). It is interesting to note that, in total, nearly half of the respondents who began the online survey did not finish. However the completion rates are not surprising given that each survey contained 85 question items, all required through forced answering, of varying difficulty, spread across six online survey pages, and taking over 11 minutes on average to finish.

Of further note in Figure 4 is the increase in the completion refusal rate that occurs across both surveys, at each of the six page breaks. This significant

![Graph](image-url)

**Notes:**
1. Vertical dotted lines indicate page breaks between the 12 survey pages.
2. The refusal rate (i.e. one minus the completion rate) difference is significantly different for the highly interesting question treatment when compared to the control survey ($p < 0.05$) from question 14 through 85 (the survey end); in 'refusal' the respondent stopped taking the survey prior to the question (and not just refused to answer the one question as all questions were required in both surveys to facilitate measurement).

**Figure 5** Highly interesting question (Treatment) reduces completion refusal rate: Hypothesis 1
completion rate variation *between* the two surveys shows the *treatment* effect (i.e. the one-stage versus two-stage highly interesting question) and is the result of introducing the second-stage portion of the same highly interesting question.

**Differences between respondents**

We used Mann-Whitney U non-parametric tests and independent sample *t*-tests with unequal variances to compare means of early and late respondents (Choi & Marden 1997; Smith & Albaum 2005); the results are consistent. We find the convention of early versus late respondent mean score testing does *not* detect the presence of non-response bias between the one- and two-stage surveys, supporting Hypothesis 2. Figure 6 shows the confidence interval plots for each variable across the two surveys. Visual inspection indicates the possibility of a pattern, with the two-stage survey group always scoring lower for Bush and higher for Ford and Wal-Mart than the one-stage survey group. However, we note the non-significance difference in the mean scores across all of the variables.

Further, visual inspection reveals that the standard deviation is consistently smaller for the two-stage survey group than the one-stage survey group. The Levene’s test for equality of variance indicates that these differences in variance are statistically significant for half the variables.

Moreover, the analysis of the statistically significant $g^2$ kurtosis statistics shows kurtosis consistently closer to the zero point in the two-stage highly interesting question survey across the variables, supporting Hypothesis 3 (see Figure 7). Likewise, the Malhita multivariate kurtosis test (DeCarlo 1997) is smaller for the two-stage highly interesting question survey than for the one-stage highly interesting question survey across the 12 descriptive variables. Thus, the two-stage highly interesting question survey contains more middle-scale (or ‘normally distributed’) answers.1 These findings suggest a difference in distribution between the two surveys not detectable by the conventional mean score practice.

**Identifying incremental respondents**

Recalling that the 12 questions used in the subgroup probability analysis are composed of four variables (i.e. Good for America, Innovative, Likeable

---

1 Readers should note that, while the one-stage highly interesting question survey responses contain more answers towards both scale ends, these should not be confused with ‘extreme responses’ in which respondents answer all questions at one scale extreme/end or the other extreme/end (see Greenleaf 1992). An evaluation of individual respondent answers across all 12 variables indicated that this is not the case.
Two-stage highly interesting questions in online marketing research

Figure 6 Descriptive confidence intervals show that conventional means tests are not capable of identifying non-response bias for one- and two-stage HIQ surveys: Hypothesis 2

Notes:
1. The figure contains the mean score and 90% confidence interval scores for each of the 12 descriptive variables for both the one-stage highly interesting survey (in black) and two-stage highly interesting question survey (in grey). Data have been centred along the mean scores of the one-stage version to permit easier comparison of the distributions of the two surveys. While the intervals are true to scale, readers should refer to the numerical scores in the figure when making assessments.
2. This graphical depiction of mean scores in Table 1 illustrates the support for Hypothesis 2: the conventional mean-score testing does not detect non-response bias differences between the control and treatment surveys.
First, cluster one comprises individuals who indicate fairly negative beliefs as to Bush, Ford and Wal-Mart. In contrast, cluster-two respondents indicate rather positive beliefs for Bush and Wal-Mart, but neutral beliefs on Ford. Cluster-three individuals indicate positive beliefs as to Ford and Wal-Mart, but negative views towards Bush. Cluster four states negative views of Wal-Mart, but very positive views of Bush and Ford. Cluster five indicates very negative views of Bush, slightly negative views of Wal-Mart, and rather positive views of Ford. Lastly, cluster six indicates very favourable views of all three subjects. All of the respondents from the two-stage survey were successfully mapped on to one of the six clusters from the one-stage survey – with the exception of the newly identified subgroup of two-stage survey respondents shown in Figure 8, who indicate a more moderate, slightly positive view across all the 12 leadership scores on the three subjects.

In detail, 12% of the respondents in the two-stage question survey did not fit into any of the six clusters of responses (across the 12 descriptive variables) from the one-stage question survey – that is, they had a probability score of less than .5 for all six clusters. Chi square distribution analysis to test kurtosis levels shows that levels were more normal in this incremental group (in the two-stage highly interesting question survey) versus the rest of the two-stage highly interesting survey respondents (as well as the one-stage highly interesting survey respondents). A detailed graphical summary (of the mean scores for the incremental respondent group) is found in Figure 8.
Two-stage highly interesting questions in online marketing research

Notes:
1. Rows represent the deviation from the centre scale (1 to 5 scales) for each of the 12 descriptive variable mean scores by cluster (i.e., the cluster mean variable score minus 2.5). The vertical axis (1 through 12 for each cluster) is the 12 descriptive variables in the study. For example, cluster 1 indicates fairly strong negative views across the four leadership variables (Good for America, Innovative, Likeable, Competent) for Bush, Ford, and Wal-Mart. In contrast, cluster 3 indicates negative views on Bush (rows 1-4), but positive views on Ford (rows 5-8) and Wal-Mart (rows 9-12).

2. The new subgroups (in light grey scale) in the two-stage highly interesting (HI) question survey are different from all of the sets of respondent patterns identified in the one-stage survey. In particular, they are closer to the middle scale values on all variables, indicating more moderate views/less interest in the survey questions.

Figure 8 Multivariate response pattern of the new subgroup in the two-stage HI question survey vs the six segments in the one-stage HI question survey
In short, the two-stage highly interesting question induced responses from an incremental group of approximately 12% of respondents, whose answer patterns across the several descriptive variables are more scale-centric (i.e. closer to the centre of the scales) than any of the cluster patterns identified in the one-stage highly interesting question survey (further support for the incremental response rate from two-stage highly interesting questions proposed in Hypothesis 1).

Discussion

Groves (1987, p. 166) poses the question ‘Why isn’t all the knowledge about survey errors used in practice?’ In regards to using highly interesting questions, part of the answer may be the relative difficulty of implementation. Offering a monetary incentive or raffle is an easy recipe to follow, and economic rationality causes us to believe that participants are more likely to respond to monetary incentives than highly interesting questions. Further, the perceived environmental shaping of a particular highly interesting question is bound to historical events and could become outdated or even decrease in effect with frequency of exposure. Also, creating and testing highly interesting questions requires more pre-survey effort than does the allocation of economic incentives.

The results of this study provide evidence that economic incentives (extrinsic rewards) and survey salience (internal rewards) are somewhat non-compensatory for upwards of 12% of respondents. In this light, researchers tend to make poor choices when they settle for the easy solution and ignore answers that may contradict common conventions. Indeed, a lack of differences in early and late respondents is not a satisfactory substitute for best practices. Wright and Armstrong (2008) argue that such testing is incorrect and a misinterpretation of Armstrong and Overton (1977).

This study provides tri-part empirical support for such assertions. First, there are no apparent mean score differences between the two survey versions (see Table 2 and Figure 5). Second, however, there is an 8% difference in completion refusal rate between the two surveys (see Figure 4). Third, moreover, subgroup probability modelling indicates that this 8% reduction most likely comprises individuals with more ‘multivariate middle of the road’ opinions (see Figure 7) – that is, these additional respondents in the two-stage highly interesting question survey indicate views that are different from those answering the one-stage highly interesting question survey, as observed through kurtosis summary statistics (Figure 6), but not mean score summary statistics (Figure 5).
Two-stage highly interesting questions in online marketing research

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Highly interesting question location and variable measures: means and kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One-stage</td>
</tr>
<tr>
<td>N (completed)</td>
<td></td>
</tr>
<tr>
<td>George W. Bush</td>
<td></td>
</tr>
<tr>
<td>Good for America</td>
<td>2.6</td>
</tr>
<tr>
<td>Innovative</td>
<td>2.5</td>
</tr>
<tr>
<td>Likeable</td>
<td>2.7</td>
</tr>
<tr>
<td>Competent</td>
<td>2.6</td>
</tr>
<tr>
<td>Ford</td>
<td></td>
</tr>
<tr>
<td>Good for America</td>
<td>3.2</td>
</tr>
<tr>
<td>Innovative</td>
<td>3.2</td>
</tr>
<tr>
<td>Likeable</td>
<td>3.1</td>
</tr>
<tr>
<td>Competent</td>
<td>3.2</td>
</tr>
<tr>
<td>Wal-Mart</td>
<td></td>
</tr>
<tr>
<td>Good for America</td>
<td>3.2</td>
</tr>
<tr>
<td>Innovative</td>
<td>3.2</td>
</tr>
<tr>
<td>Likeable</td>
<td>3.2</td>
</tr>
<tr>
<td>Competent</td>
<td>3.3*</td>
</tr>
</tbody>
</table>

Note: *p < 0.10, **p < 0.05, ***p < 0.01.

The 12 descriptive variables (e.g. Bush Good for America, Wal-Mart Competent) were compiled using attributes from the 'Big 5 Scale' (see, e.g., Digman 1990). Eleven of 12 tests fail to reject the potential of no difference in mean scores between the two concurrent surveys (supporting H2) at either a 90% or 95% confidence interval. In contrast, the standard deviation is consistently smaller supporting Hypothesis 3a, and the kurtosis score is less platykurtic (i.e. more normal) across the variables, supporting Hypothesis 3b.

These findings alone do not assist future research. If, as we found, mean score comparisons are not sensitive enough to identify non-response bias, what should practitioners and researchers do instead to identify it?

One approach is to require a particular minimum response rate. However, higher response rates do not necessarily equate to lower non-response bias (e.g. Curtin et al. 2000; Keeter et al. 2000; Groves & Peytcheva 2008). Non-response bias is the interaction of response rate and the differences between non-respondent and respondent survey scores. This for example might occur when the survey topic is a factor in the participation decision (Groves et al. 2004). In other cases, lower response rates occur because researchers are interested in the opinions of individuals who simply have many time demands or are weakly integrated into the sampling frame community. Thus a lower response rate, by itself, is not indicative of non-response bias. In fact, the opposite is also possible. A study about the affects of HEPA laws on medical office practices might be more generalisable based on a sample achieving a 3% response rate of physicians rather than a 70% response rate of undergraduate students, or
even a 50% response rate of a general populace survey panel, as bias is relative to the theoretical target population.

A second approach, as discussed by Groves (2006), is to sample non-respondents and compare their distribution of answers to the sample (e.g. Curtin et al. 2000, 2005; Lin & Schaeffer 1995; Wilcox 1977). However, this increases cost or required resources, and can be difficult – if not impossible – in some circumstances. We note, for instance, in the case of online surveys panels, the researcher usually does not have direct access to the individuals on the list maintained by the panel provider. Thus, resampling of non-respondents is not possible – as was the case in the study here. Further, many respondents may not be interested in participating in a resampling (e.g. fatigue). Or, in the case of mail or telephone surveys, individuals change locations, positions and companies.

This paper focuses on a third approach (which could be joined with either of the previous two approaches; the approach involves engaging in additional statistical testing beyond mean scores, such as standard deviation and kurtosis, to compare waves of respondents and identify the potential presence of non-response bias. We show that the inclusion of a two-stage highly interesting question might assist in decreasing non-response bias. This approach neither requires additional funding nor poses tracking problems. We recommend that practitioners consider how they might implement two-stage highly interesting questions in their surveys. For scholars, we recommend additional research that investigates the impact of two-stage highly interesting questions across other survey formats (e.g. telephone, cell phone). In short, while the results of this paper provide empirical, inductive support for the third option, this does not end the discussion – rather, it is intended to promote further inquiry.

Opinion researchers need to evaluate whether more responses are always a good thing – that is, if completed responses could be obtained from everyone, would researchers really want the additional answers? Researchers obviously want answers that provide additional value. However, Wilcox (1977) empirically demonstrates that increases in response rate can, under certain conditions, result in biased sample estimates. Therefore, research is needed that investigates (1) whether additional responses, due to salience or incentives, are informing and (2) to what extent the uninformed, unformed or uncaring responses add value or simply create noise.

As a final note, future research is needed into the optimisation of highly interesting question topics to increase response rates. Do more optimal types of highly interesting questions exist that involve current events, politics, business, sports, personalities, humour, etc? Likewise, can dynamic questioning identify
Two-stage highly interesting questions in online marketing research

topics of most interest to the respondent and that increase the likelihood of producing a response? Alternatively, in what situations could researchers engineer effective highly interesting questions by reframing the extant survey topic in an additional question as a matter of policy intervention? When would such questions obtain more responses from a more generalisable sample? When do they increase versus decrease different forms of survey error? Or, to what extent does interest level affect bivariate relationships in research, from effect size to statistical significance to directionality?

Historically, non-response bias has been used as a bar of rejection in opinion research across the social sciences rather than a measure of the generalisability of findings. Indeed, most survey research articles probably contain some non-response bias if a societal aggregate of individuals is used as the baseline. Indeed, few researchers ever succeed in avoiding the problem of non-response (Ellis et al. 1970, p. 103). Measurement theory and application should complement the methodological theory of investigation. Kurtosis, as a descriptive measure of potential non-response bias, provides an initial evaluation of the distribution of the topic matter of interest. As such, it is not a threshold (like mean scores); rather it reinforces theory, enhancing the appropriateness of the sample design.

References


Two-stage highly interesting questions in online marketing research


About the authors

Jared M. Hansen is Assistant Professor of Marketing at the University of North Carolina at Charlotte. He received his PhD from Texas Tech University, his MBA from Brigham Young University and his BS from Brigham Young University. His research examines managerial and societal implications of issues as they relate to consumer decision making and social good, retailing/marketing strategy and theory construction/measurement.

Scott M. Smith is President of Qualtrics Labs, Inc. and Professor Emeritus at Brigham Young University. He received his PhD from the Pennsylvania State University, MBA from Michigan State University and BS from Brigham Young University. He is the author or co-author of numerous books and articles in referred journals and conference proceedings. His writings deal with issues in Internet marketing, online survey research methods and cross-cultural research.

Address correspondence to: Scott M. Smith, Professor Emeritus, Marriott School of Management, Brigham Young University, Provo, Utah, USA.

Email: smsmith@byu.edu