ABSTRACT

Title of Dissertation: BEYOND RESPONSE RATES: THE EFFECT OF PREPAID INCENTIVES ON MEASUREMENT ERROR

Rebecca Medway, Doctor of Philosophy, 2012

Dissertation directed by: Dr. Roger Tourangeau

As response rates continue to decline, survey researchers increasingly offer incentives as a way to motivate sample members to take part in their surveys. Extensive prior research demonstrates that prepaid incentives are an effective tool for doing so. If prepaid incentives influence behavior at the stage of deciding whether or not to participate, they also may alter the way that respondents behave while completing surveys. Nevertheless, most research has focused narrowly on the effect that incentives have on response rates. Survey researchers should have a better empirical basis for assessing the potential tradeoffs associated with the higher responses rates yielded by prepaid incentives.

This dissertation describes the results of three studies aimed at expanding our understanding of the impact of prepaid incentives on measurement error. The first study explored the effect that a $5 prepaid cash incentive had on twelve indicators of respondent effort in a national telephone survey. The incentive led to significant reductions in item nonresponse and interview length. However, it had little effect on the other indicators, such as response order effects and responses to open-ended items. The second study evaluated the effect that a $5 prepaid cash incentive had on responses to

sensitive questions in a mail survey of registered voters. The incentive resulted in a significant increase in the proportion of highly undesirable attitudes and behaviors to which respondents admitted and had no effect on responses to less sensitive items. While the incentive led to a general pattern of reduced nonresponse bias and increased measurement bias for the three voting items where administrative data was available for the full sample, these effects generally were not significant. The third study tested for measurement invariance in incentive and control group responses to four multi-item scales from three recent surveys that included prepaid incentive experiments. There was no evidence of differential item functioning; however, full metric invariance could not be established for one of the scales.

Generally, these results suggest that prepaid incentives had minimal impact on measurement error. Thus, these findings should be reassuring for survey researchers considering the use of prepaid incentives to increase response rates.

BEYOND RESPONSE RATES: THE EFFECT OF PREPAID INCENTIVES ON MEASUREMENT ERROR

By

Rebecca Medway

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2012

Advisory Committee:
Dr. Roger Tourangeau, Chair
Professor Frauke Kreuter
Professor Stanley Presser
Professor John Robinson
Professor Eleanor Singer

ACKNOWLEDGMENTS

The research presented in this dissertation was made possible through the support of several organizations. I am indebted to JPSM for allowing me to include my research in the 2011 JPSM Practicum survey. The results of this study became a focal point of both Chapters 2 and 4 of my dissertation. Additionally, I am grateful to the U.S. Census Bureau for awarding me a Dissertation Fellowship, which allowed me to conduct the mail survey featured in Chapters 3 and 4 of this dissertation. I also am grateful to the Survey Research Center at the University of Michigan for providing me access to data from the Survey of Consumers.

There also are several individuals whose help was invaluable in creating this dissertation. First, I'd like to thank the members of my committee, particularly my Chair Roger Tourangeau, for their helpful feedback and unwavering support over the past two years. And finally, thank you to my parents, Marc and Cathy: for nurturing a love of learning from my earliest days and for always pushing me to be the best version of myself – both in and out of the classroom.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INCENTIVES AND SURVEY RESEARCH

1.1 INTRODUCTION

As survey response rates continue to decline, incentives are increasingly used as a way to motivate sample members to participate (Cantor, O'Hare, & O'Connor, 2008; Kulka, Eyerman, & McNeeley, 2005; Singer, Van Hoewyk, & Maher, 2000). Extensive research shows that incentives can increase response rates (e.g., Church, 1993; Hopkins & Gullickson, 1992; Singer, Van Hoewyk, Gebler, Raghunathan, & McGonagle, 1999; Trussell & Lavrakas, 2004); they clearly convince some sample members to participate who otherwise would not have done so. If they influence behavior at the stage of deciding whether or not to participate, it is reasonable to believe that incentives also may alter the way that respondents act during the survey interview.

Thus, it is important to determine whether the use of incentives influences the magnitude of measurement error in survey estimates. Nevertheless, as pointed out by Singer and Ye (forthcoming), the majority of incentives research has focused narrowly on the effect that incentives have on response rates. Groves (2008) voices similar concerns and urges researchers to, "re-conceptualize the focus away from response rates". Likewise, Cantor et al. (2008) speak to the need to improve our understanding of the impact that incentives have on data quality.

Incentives conceivably could lead to an increase or a decrease in measurement error. On one hand, they could reduce measurement error if they create a sense of obligation to the researcher that causes respondents to make greater effort and provide more thorough, thoughtful responses to questions. Such a result would be reassuring to

survey practitioners who are enticed by the promise of higher response rates but lack sufficient empirical evidence of other benefits to justify the costs that can be associated with incentives. On the other hand, incentives could increase measurement error if they convince otherwise uninterested sample members to participate who lack intrinsic motivation to do so. As Brehm (1994) argues, "If we happen to get 100 percent of our respondents, but they all told us lies, we get 100 percent garbage" (p. 59). Survey practitioners should have a better empirical basis for assessing the potential tradeoffs associated with the higher responses rates yielded by prepaid incentives.

This dissertation aims to expand our understanding of the impact of prepaid incentives on measurement error. In this chapter, I begin by reviewing the existing literature assessing the effect of incentives on both nonresponse and measurement error. In the three analytical chapters that follow, I address the following questions in turn: Do incentives affect the level of effort that respondents put into completing surveys? Do they influence self-presentation concerns, thereby altering responses to sensitive questions? Finally, does measurement invariance exist between responses received with an incentive and those received without one?

1.2 INCENTIVES AND NONRESPONSE

1.2.1 Declining Survey Response Rates

Survey response rates have declined considerably over the past several decades (Brick & Williams, forthcoming; Steeh, Kirgis, Cannon, & DeWitt, 2001). For example, the response rate for the National Immunization Survey decreased by fourteen percentage points between 1995 and 2004 (Battaglia et al., 2008), while the response rates for the Consumer Expenditure Diary Survey and the National Health Interview Survey declined by twelve and eight percentage points, respectively, in the 1990s (Atrostic, Bates, Burt & Silberstein, 2001). De Leeuw and de Heer (2002) demonstrate that this is an international phenomenon; reviewing a multi-national sample of household surveys, they report that response rates have decreased by an average of half a percentage point per year over the past twenty years. Furthermore, the speed of this decline may be increasing; the response rate for the Survey of Consumer Attitudes decreased by one and a half percentage points per year from 1996 to 2003 – double the average annual decline observed from 1979 to 1996 (Curtin, Presser, & Singer, 2005).

Low response rates can be problematic for several reasons. First, although the response rate is not always a good predictor of nonresponse bias (Groves, 2006), lower response rates may increase the potential for nonresponse bias. Nonresponse bias is a function of both the response rate and the difference between respondents and nonrespondents on survey estimates; if those individuals who respond are not representative of the larger sample on the variables of interest, the estimates will be biased (Groves & Couper, 1998). Furthermore, low response rates may increase survey costs, as they mean that larger initial samples are required to attain the number of

respondents necessary to achieve desired levels of precision in survey estimates (Groves, Dillman, Eltinge, & Little, 2002).

Survey nonresponse generally can be broken into two major components: inability to reach sample members ("noncontacts") and failure to persuade them to complete the survey once they have been contacted ("refusals"). Research on surveys in modes where we can more easily disentangle these two components, such as face-to-face and telephone, repeatedly demonstrates that refusals account for a considerably larger proportion of nonresponse than do noncontacts (Brick & Williams, forthcoming; Curtin et al., 2005; Smith, 1995). Typical reasons provided for refusing include being too busy, not being interested in the survey topic, privacy concerns (such as not wanting to share personal information with a stranger), or negative reactions to aspects of the survey (such as its length) (Brehm, 1993; Bates, Dalhamer & Singer, 2008; Couper, Singer, Conrad, & Groves, 2008). Incentives may be an effective tool for reducing some of these refusals – either as an expression of gratitude for respondents' time or as a way of overcoming a lack of interest in the survey topic. In fact, Couper et al. (2008) found that, following altruistic desires to be helpful or to influence policy, receiving money was one of the most common reasons provided for agreeing to respond to a (hypothetical) survey request.

The results of several studies suggest that incentives' effect on the response rate is largely a function of reduced refusal rates (Holbrook, Krosnick, & Pfent, 2008; Shettle & Mooney, 1999; Tourangeau, Groves, & Redline, 2010; Willimack, Schuman, Pennell, & Lepkowski, 1995). However, many other studies have not disentangled the effect of incentives on noncontact from their effect on refusals (Singer & Ye, forthcoming) –

possibly because such a large proportion of the existing experiments are part of mail surveys, where it can be difficult to determine whether nonresponse is caused by lack of contact or lack of willingness to participate.

1.2.2 Effect of Incentives on Response Rates

Survey practitioners are searching continually for ways to combat declining response rates. Several tools, such as pre-notification, multiple follow-up contacts, and incentives, have proven effective and have become part of common survey practice. For example, in a meta-analysis of 251 mail surveys, Edwards and colleagues found that offering cash incentives doubled the odds of response, while pre-notification and follow-up contacts each multiplied the odds of response by about 1.5 (Edwards et al., 2002).

Numerous experimental studies have demonstrated that incentives are an effective tool for increasing survey response rates (e.g., James & Bolstein, 1990; Shettle & Mooney, 1999; Petrolia & Bhattacharjee, 2009; Yammarino, Skinner, & Childers, 1991; Yu & Cooper, 1983). Several meta-analyses have shown that the successfulness of incentives spans all survey modes. For example, Church (1993) found that offering an incentive in mail surveys increases the response rate by an average of 13 percentage points. Similarly, incentives multiply the odds of response to Internet surveys by 1.3 on average (Göritz, 2006). Finally, a meta-analysis of both prepaid and promised incentive experiments in interviewer-administered surveys confirmed that incentives have a positive, but smaller, impact on response rates in these types of surveys as well; in these experiments, each dollar that was given to respondents increased the response rate by about one-third of a percentage point on average (Singer et al., 1999).

Certain types of incentives have proven to be more effective than others. Prepaid incentives tend to be more successful than promised ones contingent on completion of the survey (Armstrong, 1975; Berk, Mathiowetz, Ward, & White, 1987; Church, 1993; James & Bolstein, 1992; Petrolia & Bhattacharjee, 2009), and monetary incentives tend to be more effective than non-monetary ones (Hansen, 1980; Petrolia & Bhattacharjee, 2009; Warriner, Goyder, Gjertsen, Hohner, & McSpurren, 1996). As these findings imply, prepaid monetary incentives generally have the greatest impact on response rates. Two separate meta-analyses of incentive experiments in mail surveys both concluded that prepaid cash incentives increase mail survey response rates by 19 percent points on average (Church, 1993; Hopkins & Gullickson, 1992). Replicating Singer et al.'s (1999) finding that incentives have a smaller impact in interviewer-administered surveys, Cantor et al. (2008) found that prepaid cash incentives of up to $10 led to a median increase of six percentage points in RDD surveys.

While some studies observe a linear relationship between the value of the incentive and the increase in the response rate (e.g., Church, 1993; Trussell & Lavrakas, 2004; Yu & Cooper, 1983), others conclude that increases in the incentive value may have diminishing influence on the response rate (Cantor et al., 2008; Fox, Crask, & Kim, 1988; James & Bolstein, 1992). Finally, although this dissertation generally focuses on the use of incentives at initial contact in cross-sectional surveys, incentives also have proven effective in other contexts. For example, incentives may reduce attrition in longitudinal studies (Creighton, King, & Martin, 2007; Goetz, Tyler, & Cook, 1984), and they may be an effective tool for refusal conversion (Brick, Montaquila, Hagedorn, Roth, & Chapman, 2005).

1.2.3 Theoretical Explanations for the Effectiveness of Incentives

Multiple theories of survey response provide potential explanations for incentives' success at increasing response rates. For example, utility theory suggests that individuals weigh the costs and benefits of completing a task and will take action when the benefits of doing so exceed the costs (Groves & Couper, 1998). Offering an incentive is one way that researchers can make the perceived benefits of taking part in survey research greater than the perceived costs. Under such a framework, respondents may see the incentive as payment or reimbursement for their time and effort (Biner & Kidd, 1994). Conceptualizing the incentive as an economic exchange helps to explain why larger incentives have at times been found to be more effective than smaller ones (e.g., Trussell & Lavrakas, 2004).

Other researchers have suggested that the effectiveness of incentives is not due to an economic exchange but a social one. Under social exchange theory (Blau, 1964; Homans, 1961), rewards and costs remain important decision-making factors, and individuals still choose to take action only when they feel it is in their self-interest to do so. However, social exchange is different from economic exchange in two main ways (Dillman, Smyth, & Christian, 2009). First, in social exchange, the definition of rewards and costs are more flexible; namely, the rewards do not have to be monetary. Second, the importance of trust is much greater in social exchanges. Social exchanges typically are not bound by contracts, and so individuals have to trust that the other party will provide a reward in the future that will be worth whatever cost they must bear.

Actors in such exchanges are able to trust one another due to several rules and norms of exchange by which they can assume the other party will abide. One of the

central rules of social exchange is the norm of reciprocity; this rule suggests that when an individual takes an action that benefits you, you are expected to respond in kind (Gouldner, 1960). Incentives can be seen as a benefit that the survey sponsor provides to the sample member; when sample members receive an incentive they may feel obligated to return the kindness by responding to the survey. This may explain the effectiveness of prepaid incentives (Dillman et al., 2009). However, the mixed success of promised incentives suggests that sample members do not trust survey researchers enough to incur the costs of participation without having received their reward in advance.

Other researchers have suggested a related explanation for the effectiveness of prepaid cash incentives, based on cognitive dissonance theory (Festinger, 1957). According to this theory, once respondents have received a prepaid incentive, the idea of keeping it without completing the survey creates a feeling of dissonance (Furse & Stewart, 1982). Sample members have two options for resolving this unpleasant feeling. The first is to dispose of the incentive; however, Furse and Stewart (1982) argue that most people will not choose this option because throwing money away also makes them feel uncomfortable, and because sending the money back to the researcher may involve almost as much effort as simply completing the survey. Therefore, most people will choose the second option – participating in the survey.

Finally, leverage-saliency theory suggests that the impact of various design features on the participation decision differs across sample members (Groves, Singer, & Corning, 2000). According to this theory, the influence of each feature on an individual's decision to respond depends on three factors: (1) how important the feature is to the sample member (leverage), (2) whether the sample member sees this as a positive or

negative feature (valence), and (3) the degree to which the feature is highlighted in the survey request (salience). For example, some sample members may choose to respond because they are interested in the survey topic described in the survey cover letter. Other sample members may lack such an interest but may be convinced to participate by a cash incentive included in the envelope. Thus, incentives may convince certain sample members to respond who are not drawn to other survey features such as the topic, and they may have little or no effect on other sample members' willingness to participate (e.g., Baumgartner & Rathbun, 1996).

1.2.4 Effect of Incentives on Sample Composition and Nonresponse Error

Several experimental studies in both mail and interviewer-administered modes have found that incentives, whether prepaid or promised, do not have much of an effect on sample composition (e.g., Brick et al., 2005; Cantor et al., 2008; Furse & Stewart, 1982; Goetz et al., 1984; James & Bolstein, 1990; Shettle & Mooney, 1999; Warriner et al., 1996; Willimack et al., 1995). However, the results of other experiments suggest that incentives can have two types of effects on sample composition. First, incentives may improve representation of traditionally underrepresented groups, such as young people (Dillman, 1996; Miller, 1996; Storms & Loosveldt, 2004), minorities (Berlin et al., 1992; Mack, Huggins, Keathley, & Sundukchi, 1998), and those with lower incomes (Mack et al., 1998) or less education (Berlin et al., 1992; Nederhof, 1983; Petrolia & Bhattacharjee, 2009).

Second, the use of incentives may alter the characteristics of the respondent pool along dimensions other than the typical demographic variables measured in surveys. For example, as leverage-saliency theory might predict, incentives may help attract

respondents who are less interested in the survey topic (Baumgartner & Rathbun, 1996; Coogan & Rosenberg, 2004; Petrolia & Bhattacharjee, 2009). However, in a series of prepaid cash incentive experiments embedded in mail and telephone surveys, Groves and colleagues found only mixed support for this hypothesis (Groves et al., 2006; Groves, Presser, & Dipko, 2004). Additionally, Moyer and Brown (2008) actually found the reverse effect: promising a cash incentive for completing the National Cancer Institute's telephone-administered Health Information National Trends Survey (HINTS) significantly increased the proportion of respondents who had had cancer.

The use of incentives also may reduce the proportion of respondents with certain personality traits or values, such as altruism or selflessness, due to an influx of more selfish respondents. Altruistic or selfless sample members are likely to respond to surveys even without an incentive, while incentives may serve as a motivating factor for sample members low in these traits (Storms & Loosveldt, 2004). For example, in a mail follow-up to the Detroit Area Study (DAS), Groves et al. (2000) found that a $5 prepaid cash incentive had a significantly greater impact on the response rate among DAS respondents who had reported low levels of community involvement than it did among those who had reported being more involved. Medway and colleagues found that offering a $5 prepaid incentive increased the proportion of respondents to a mail survey who had not volunteered in the past year – although this same effect was not found in an equivalent experiment conducted as part of a telephone survey (Medway, Tourangeau, Viera, Turner, & Marsh, 2011).

To the extent that incentives improve representation of groups that are underrepresented when incentives are not used, they may lead to a reduction in

nonresponse error. This seems particularly likely in cases where incentives improve representation of individuals who lack interest in the survey topic. For example, Tourangeau et al. (2010) found that offering a prepaid cash incentive of $5 improved representation of nonvoters and reduced the nonresponse bias in reports of voting behavior in two recent elections by about six percentage points – although these differences did not reach statistical significance. However, improved representation of demographic groups that traditionally are underrepresented in surveys will reduce nonresponse bias only if these groups also differ from better-represented groups on key survey variables. For example, in an experiment that assigned sample members to receive either $5 cash or a pass to a local park, Ryu, Couper, and Marans (2005) found that the two types of incentives resulted in differences in respondents' education level, marital status, and work status; however, they did not find differences in the other response distributions for the two groups. Finally, in their meta-analysis of nonresponse bias analysis studies, Groves and Peytcheva (2008) reported that, overall, the use of an incentive did not have a significant impact on the magnitude of nonresponse bias – though very few of the studies included in their analysis made use of incentives.

1.3 INCENTIVES AND MEASUREMENT ERROR

Measurement error is any inaccuracy in survey responses that is due to the process of measurement; this type of error can be differentiated from nonresponse error, discussed earlier, which arises from the failure to get some sample members to respond in the first place. Measurement error exists when the measured value in a survey differs from the corresponding unobserved "true" value (Borhnstedt, 2010), although it may be difficult, or even impossible, for the researcher to know this true value. Several potential sources of measurement error have been identified in the literature, including the interviewer, the respondent, and features of the survey design, such as mode of administration or question wording (Groves, 1989). Offering an incentive is an additional design decision that could have repercussions for the magnitude of measurement error in the resulting estimates. However, this possibility has received relatively little attention in the literature as compared to the effect of incentives on nonresponse.

1.3.1 Theory-Based Expectations for Effect of Incentives on Measurement Error

Incentives conceivably have the potential to either increase or decrease measurement error through their influence on respondent behavior. The theories used to explain why incentives convince sample members to respond have conflicting implications for the effect of incentives on the quality of the answers provided during the interview. For example, according to social exchange theory, offering prepaid incentives is potentially the first step toward building a positive relationship between the researcher and the respondent; giving sample members a reward before receiving their responses implies that the researcher trusts and respects them. If respondents are motivated by a sense of community with the researchers, they may feel more comfortable while completing the survey, and, as a result, they may put forth more effort than they would

have otherwise. They also may be more willing to respond honestly to questions that are typically subject to social desirability biases. For example, a review of 74 incentive experiments in laboratory studies suggested that self-presentation concerns were reduced among the participations who had received incentives (Camerer & Hogarth, 1999). However, this feeling of having a positive relationship with the researcher could also lead respondents to focus too heavily on pleasing the researcher; as a result, respondents may provide more positive ratings, either generally across all items or specifically for items referring to the survey sponsor, than they would have otherwise.

Offering respondents incentives also could affect their motivations for completing the survey; in particular, it may lead them to focus on extrinsic motivations instead of intrinsic ones. For example, according to social exchange theory, incentives may create a sense of obligation toward the researcher, and this feeling may be what motivates sample members to respond. Another possibility, as suggested by leverage-saliency theory, is that incentives may convince otherwise uninterested sample members to respond. In both cases, respondents are focused on an extrinsic motivation, as opposed to an intrinsic one. It seems reasonable that people who are motivated by extrinsic factors such as monetary rewards may put forth less effort than those who are focused on intrinsic ones, such as interest in the survey topic or enjoyment of sharing one's opinions. Research on the importance of intrinsic motivation to academic success has supported this assumption (e.g., Bolkan, Goodboy, & Griffin, 2011; Fransson, 1977).

Research on the quality of responses provided by reluctant respondents, who can be assumed to have low levels of motivation to participate, suggests that such respondents sometimes provide lower quality data than do more eager respondents

(Cannell and Fowler, 1963; Triplett, Blair, Hamilton, & Kang, 1996; Fricker & Tourangeau, 2010); however, other studies have failed to find a clear relationship between reluctance and data quality (Kaminska, McCutcheon, & Billiet, 2010; Yan, Tourangeau, & Arens, 2004).

A final possibility is that, once sample members have committed to taking part in the survey, the incentive has little to no further impact on their behavior. Social exchange theory suggests that sample members are driven to respond by a sense of obligation to the researcher, while cognitive dissonance theory suggests they are driven by the desire to avoid the dissonance associated with refusal once they have accepted the incentive. If agreeing to participate in the survey satisfies these needs, then any further behaviors taken during data collection may not be influenced by the fact that the respondent has received an incentive. In support of the non-importance of incentives on respondent behavior while completing the survey, Camerer and Hogarth's (1999) review of incentive experiments in laboratory studies found that incentives typically do not affect performance in such studies.

1.3.2 Comparison of Incentive and Control Group Response Distributions

In the incentives literature, the presence of measurement error is typically assessed in one of three ways. The first is to compare the response distributions of two or more groups of respondents who have been randomly assigned to different experimental conditions. Differences between the groups' responses suggest that there may be a greater amount of error in one of the groups. However, it can be difficult to know whether these differences are caused by a change in who responds (nonresponse error) or by a change in how they respond (measurement error). Furthermore, in the absence of some gold

standard to which we can compare the survey responses, we cannot easily tell which of the groups exhibits more error.

There is some evidence that offering incentives can affect survey response distributions. Generally, these differences have been observed for attitudinal items in studies that have offered prepaid cash incentives - suggesting that incentives can lead to more positive survey responses. For example, respondents who received a prepaid cash incentive in a mail survey offered more positive comments about the sponsor in open-ended items than did those who did not receive an incentive; the researchers argue this occurred because receiving the incentive led to increased favorability toward the sponsor (James & Bolstein, 1990). In the telephone-administered Survey of Consumers, offering a $5 prepaid cash incentive had a significant effect on responses to four of seventeen key attitudinal questions; the authors suggest this happened because receiving the incentive put the respondents in a good mood (Singer et al., 2000). Brehm (1994) also found that offering a prepaid cash incentive led to more positive responses to several political attitude questions in a telephone survey. The use of prepaid cash incentives led to greater reported levels of concern about social issues for six of ten items in a mail survey, though the incentives did not increase respondents' willingness to pay to improve the condition of these social issues (Wheeler, Lazo, Heberling, Fisher, & Epp, 1997). Finally, respondents who had been promised a $1 reduction in their hotel rate in exchange for completing a questionnaire were less likely to provide negative comments about their stay, as compared to a control group (Trice, 1984).

I am aware of only two studies where incentives were found to affect response distributions to non-demographic factual items. In these cases, providing prepaid cash

incentives resulted in lower estimates of community involvement (Groves et al., 2000) and volunteering (Medway et al., 2011). However, it is impossible to know whether these differences were caused by changes in sample composition (those who volunteer their time for community activities also may be the type of people who are willing to do surveys without incentives, while those who do not do so may require an incentive to motivate them to take part in survey research) or by an increased obligation to be honest about not taking part in these socially desirable behaviors.

Several other studies have concluded that incentives do not affect response distributions. Offering a prepaid cash incentive led to significantly different responses for only five percent of the questions in a mail study (Shettle & Mooney, 1999). Similarly, overall, James and Bolstein (1990) did not find significant differences in the response distributions of 28 closed questions in a mail survey when prepaid cash incentives were offered. Offering a prepaid non-monetary incentive did not have a significant effect on responses to ten items in the face-to-face DAS (Willimack et al., 1995). Finally, offering a contingent cash incentive between $10 and $40 did not affect response distributions in two government-sponsored face-to-face studies on substance abuse (the National Survey on Drug Use & Health (NSDUH) and the Alcohol and Drug Services Study (ADSS)) (Eyerman, Bowman, Butler, & Wright, 2005; Krenzke, Mohadjer, Ritter, & Gadzuk, 2005).

It is not clear why incentives affect response distributions in some surveys and not in others. One reason may be that researchers have been inconsistent across studies in their selection of items to analyze. For example, some studies focus only on responses to the key survey questions (e.g., Curtin, Singer, & Presser, 2007; Singer et al., 2000), while

others consider all of the survey items as one large group (e.g., Shettle & Mooney, 1999). It is difficult to know whether restricting the analysis (or not doing so) is what led to these divergent results. Moving forward, the literature would benefit from a more systematic examination of the importance of various item characteristics. For example, does it matter if the item is a "key" measure that is directly related to the stated survey topic? Are attitude questions more likely to be affected by incentives than factual ones? Does the sensitivity of the item matter? How about placement in the questionnaire?

Many other recent incentive experiments fail to discuss the potential effect of incentives on response distributions. In others, the possibility of an effect is mentioned but quickly dismissed without analyzing the data; this decision is, at times, based on the results of a handful of older studies that found offering incentives did not affect survey responses. However, these older studies exhibit features that prevent their results from generalizing to all surveys offering incentives. For example, several of them used very specialized, highly educated populations and surveyed them about topics that were of specific interest to them (Goodstadt, Chung, Kronitz, & Cook, 1977; Hansen, 1980; Mizes, Fleece, & Roos, 1984). Furthermore, several of these studies differed from more recent studies in that they were able to achieve response rates of over 60%, even for the groups that did not receive an incentive (Goodstadt et al., 1977; Mizes et al., 1984; Nederhof, 1983).

1.3.3 Comparison of Survey Responses to Validation Data

A weakness of comparing response distributions is that, even when differences are observed between incentive and control group responses, it often is impossible to tell which group's responses exhibit less error. A second approach, which overcomes this

limitation, is to compare survey responses to validation data – often administrative records. In incentives research, this means that the relative accuracy of responses provided by those who have received an incentive can be compared to that of respondents who have not received one.

This method can be challenging to implement because of the difficulty of obtaining access to administrative records; however, a fair number of studies have successfully used validation to data to demonstrate that respondents often provide inaccurate answers. For example, this method has been used to demonstrate underreporting of socially desirable behaviors, such as voting (Traugott & Katosh, 1979), or respondents' difficulty recalling certain types of events, such as their children's vaccination history (Luman, Ryman, & Sablan, 2009). This method also has been used to demonstrate the impact of design features, such as survey mode, on the accuracy of survey responses to sensitive questions (Kreuter, Presser, & Tourangeau, 2008). However, using a record check to determine the accuracy of survey responses has rarely been done in conjunction with incentive experiments; in fact, in their review of the incentives literature, Singer and Ye (forthcoming) specifically point to the lack of research investigating the impact of incentives on the validity of survey responses.

I am aware of only four incentive studies that have compared the relative accuracy of reports given by respondents who have received an incentive and those who have not received one. Two of these studies offered prepaid cash incentives. The first was a mail survey of people who had bought a major appliance at one of five stores in the Midwest; sample members were randomly assigned to receive 25 cents prepaid or no incentive. McDaniel and Rao (1980) asked respondents factual questions about their

purchase (such as model name, price paid, and date of purchase) and compared respondents' reports with store records. They found that respondents who had received the prepaid incentive committed significantly fewer errors on average than members of the control group.[1]

The second study that included a prepaid cash incentive experiment was a survey of registered voters in which one half of the sample members received $5 and the other half did not receive an incentive; respondents also were randomly assigned to either mail or telephone administration. Tourangeau et al. (2010) compared respondents' self-reports of voting status in two elections to records from the Aristotle database of registered voters. They found that, for both elections, the incentive did not significantly affect the proportion of respondents that misreported. However, the direction of the effect was the same for both items – in the 2004 election the incentive led to a ten percentage point increase in the prevalence of misreporting among those who had received an incentive, and in the 2006 election it led to a five percentage point increase in the prevalence of misreporting.

The other two studies offered incentives contingent on completion of the survey. The first of these was a mail survey of elites, such as university professors and cabinet ministers, in 60 countries; the topic of the survey was family planning and population growth (Godwin, 1979). One third of the sample was offered a promised incentive of $25, one third was offered a promised incentive of $50, and the final third was not offered an incentive. For 28 factual questions such as, "Are contraceptives available in clinics in [country]?" Godwin compared the survey responses to published records and

[1] The authors do not mention whether there were any significant differences in the sample composition of the incentive and control groups on variables such as length of time since purchase, so we cannot be certain whether the difference in response quality was driven by changes of this nature.

the responses of any other respondents from the same country. He grouped respondents into "low", "medium", and "high" accuracy groups and found that being offered an incentive significantly increased the proportion of correct responses. This was particularly true for the group that was offered $50; 50% of these respondents fell into the "high" accuracy category, as compared to 26% of those offered $25 and only 20% of those in the control group.[2]

The final instance was an incentive experiment in the Alcohol and Drug Services Study (ADSS); this study interviewed individuals who were recently discharged from substance abuse treatment facilities. In this experiment, there were three incentive conditions and one control group. Two of these incentive groups were offered either $15 or $25 contingent on completion of a face-to-face interview, and all three groups were offered $10 in return for submitting to a urine drug test. In their analysis of this experiment, Krenzke et al. (2005) utilized the $15/$10 group as the comparison group.

The researchers reported two efforts to compare respondents' self-reports with validation data. First, 20 survey responses, mostly asking about drug use, were compared to records from the treatment facility (Table 1.1). Next, respondents' self-reports of drug use in the past seven days and past 24 hours were compared to the results of urine tests (Table 1.1).[3] Overall, these results suggest that the $15 incentive led to limited improvements in accuracy; responses to four of twenty survey items were significantly more likely to be accurate as compared to facility records, and self-reports of drug use as compared to urine tests were significantly more likely to be accurate for one of six items.

[2] The author does not discuss whether these differences may have been driven by differences in sample composition between the incentive and control groups.

[3] Respondents were told that they would be subject to a drug test before they provided the self-report; therefore, respondents who may have otherwise lied about their drug use may have decided to be honest in this particular case.

Furthermore, offering a $25 contingent incentive led to significant reductions in accuracy for two of the four survey items where we saw improvements with a $15 incentive.

Table 1.1. Effect of Incentive on Accuracy of Self-Reports (Krenzke et al., 2005)

	Compared to Treatment Facility Records	**Compared to Urine Test**
$15 Contingent Incentive vs. No Contingent Incentive	Significant improvement for 4 of 20 items; Significant reduction for 1 item	Significant improvement for 1 of 6 items
$25 Contingent Incentive vs. $15 Contingent Incentive	Significant reduction for 2 of 20 items	No effect

Thus, these four studies report conflicting results. The two studies finding an effect differ from those that did not on several dimensions. First, the two studies finding an increase in accuracy were published quite a while ago (1979, 1980), while those finding no effect were published more recently (2005, 2010). Second, the two studies that found an incentive effect were both mail studies, whereas at least some of the respondents in the two studies that did not find an effect utilized an interviewer-administered mode. Finally, the studies finding an improvement in quality looked at the accuracy of non-sensitive factual questions, while the two that found no effect looked at the accuracy of somewhat sensitive topics. Because only four studies have been conducted, it is difficult to know which of these dimensions is the most important.

1.3.4 Comparison of Effort Indicators

The prevalence of measurement error also may be assessed in a third way; in this method, respondents who have received an incentive again are compared with those who have not received one. However, this method examines indirect indicators of data quality, such as missing data rates, thought to reflect respondents' level of effort. Although effort indicators are only indirect measures of data quality, respondents who put forth greater

effort also may provide responses that have less measurement error. This method has frequently been employed in mode comparisons; for example, researchers have found that telephone respondents are more likely to satisfice than are face-to-face respondents (Holbrook, Green, & Krosnick, 2003; but see Roberts, Jäckle, & Lynn, 2007) or Internet respondents (Chang & Krosnick, 2009) and that cell phone respondents are no more likely than landline ones to take cognitive shortcuts (Kennedy & Everett, 2011).

Much of the existing literature investigating the impact of incentives on respondent effort focuses on item nonresponse or on the length of responses to open-ended questions. Many of these studies have concluded that incentives do not have a significant impact on the prevalence of item nonresponse (e.g., Berk et al., 1987; Berlin et al., 1992; Cantor et al., 2008; Curtin et al., 2007; Furse & Stewart, 1982; Peck & Dresch, 1981; Shettle & Mooney, 1999). This conclusion has been reached across a variety of incentive types and a multitude of survey characteristics. For example, sending prepaid cash incentives in a mail survey of cable subscribers did not significantly affect the proportion of items that respondents skipped (James & Bolstein, 1990). Dirmaier and colleagues came to the same conclusion in a mail survey of psychotherapy patients (Dirmaier, Harfst, Koch & Schulz, 2007). Similarly, offering a non-contingent voucher in the in-person Survey of Income and Program Participation (SIPP) did not have a significant impact on the proportion of cases that were considered "mostly complete" (Davern, Rockwell, Sherrod, & Campbell, 2003). Finally, offering a contingent incentive in the National Adult Literacy Survey did not have a significant effect on the proportion of items that respondents attempted (Berlin et al., 1992).

However, several other studies have observed a reduction in item nonresponse when incentives are utilized; again, these studies have used both prepaid and promised incentives, have been conducted in a variety of modes, and have asked respondents about a wide range of topics. For example, in a mail survey of people who had bought a major appliance at one of five stores in the Midwest, sending a prepaid cash incentive of 25 cents significantly reduced the mean number items that respondents skipped (McDaniel & Rao, 1980). Similarly, sending a prepaid debit card worth $40 in the first wave of the face-to-face Consumer Expenditure Quarterly Interview Survey significantly reduced the number of items that respondents skipped in both the first wave and subsequent waves; the use of a $20 debit card also slightly reduced item nonresponse but not significantly so (Goldenberg, McGrath, & Tan, 2009). Offering a promised incentive of either $20 or $50 significantly reduced item nonresponse in an international mail survey of elites (Godwin, 1979). Finally, offering a promised incentive of $10 in a telephone survey of Chicago residents significantly reduced the number of items that respondents skipped; this decrease was driven by a reduction in the number of "don't know" responses (Goetz et al., 1984). None of the incentive experiments I found resulted in a significant overall increase in item nonresponse.

The studies listed above provided information about how the incentive affected item nonresponse across all items for all respondents; however, it is possible that the effect of the incentive was restricted to certain subgroups of respondents or particular types of items. Only a few studies have considered either of these possibilities, and those that have done so have tended to find conflicting results. For example, Singer et al. (2000) found that receiving an incentive in the Survey of Consumers led to a significant

reduction in item nonresponse for two particular subgroups – older respondents and non-Whites; however, this result was not replicated in a similar experiment conducted in a subsequent administration (Curtin et al., 2007).

The possibility that incentives affect open and closed items differently has been considered in two studies, with conflicting results. Hansen (1980) found that providing a prepaid incentive of either 25 cents or a ballpoint pen led to a significant increase in the proportion of open-ended items that were skipped but had no effect on closed items. However, McDaniel and Rao (1980) found that offering a prepaid incentive of 25 cents significantly reduced item nonresponse for both open-ended and closed items. Two face-to-face to face studies considered the possibility that the effect of the incentive on item nonresponse might differ by item sensitivity, again with conflicting results. Providing a prepaid monetary incentive of three to five pounds in the British Social Attitudes Survey reduced item nonresponse for non-sensitive questions but increased it for sensitive ones (Tzamourani and Lynn, 1999). However, in a study of recently-released clients of drug and alcohol treatment facilities, offering a promised incentive of $15 to $25 did not have a significant impact on the proportion of items respondents skipped, regardless of item sensitivity (Krenzke et al., 2005). In this same study, the researchers hypothesized that the effect of the incentive would be greater in the final section of the interview, when respondents were tired, but this prediction was not supported by the data.

The other effort-related outcome that frequently has been analyzed in incentive studies is the quality of open-ended responses, generally operationalized as the number of words or number of ideas included in the response. As with those that have looked at item nonresponse, these studies have tended to find either an improvement in quality with

an incentive or no effect. For example, in a telephone follow-up to the National Election Studies, respondents who had received a prepaid incentive of either $1 or a pen provided significantly more ideas on average in response to two open-ended questions (Brehm, 1994). Similarly, respondents to a telephone survey who had been promised $10 provided significantly more words on average in response to open items (Goetz et al., 1984). In a contradictory finding, Hansen (1980) found that mail respondents who were given a prepaid incentive of either 25 cents or a pen provided significantly fewer words on average; coders also rated the incentive groups' responses to be of lower quality on average.

Interestingly, several studies that have offered more than one value of monetary incentive have found that only the larger amount has resulted in improved response quality. For example, in the British Social Attitudes Survey, respondents who were given five pounds provided significantly longer responses to open-ended items as compared to a control group – but receiving three pounds did not have a significant effect on response length (Tzamourani & Lynn, 1999). James and Bolstein (1990) conducted an incentive experiment as part of a mail survey in which respondents were given prepaid cash incentives of either 25 cents, 50 cents, $1, or $2. They found that only those respondents who had received at least 50 cents wrote significantly more words than the control group for an open-ended question. For a short-answer question where respondents were given space to write up to four comments, they also found that only those respondents who had received at least $1 provided a significantly greater number of comments. In a mail survey of elites, respondents who were promised $50 provided significantly more

detailed responses to an open item, but there was not a significant improvement in quality among respondents who were promised $25 (Godwin, 1979).

It is rare for incentive studies to have considered data quality indicators beyond item nonresponse and responses to open-ended items; again, the studies that have done so have generally found that incentives either improve effort or have no effect. For example, the number of events recorded in a diary increased when an incentive was provided (Lynn & Sturgis, 1997). Receiving a prepaid voucher worth either $10 or $20 did not have a significant effect on the number of imputations or edits required for 40 items in the SIPP (Davern et al., 2003). In two other recent surveys, prepaid cash incentives did not affect the number of responses selected for a check-all-that-apply item, the proportion of respondents who provided at least one pair of inconsistent responses, or the proportion of respondents who provided round numerical responses (Medway et al., 2011).

Finally, a few studies have found that respondents who have received incentives have been more willing to submit to requests that imply potential additional burden or may raise privacy concerns; for example, respondents who received prepaid incentives were more likely to provide additional contact information (Shettle & Mooney, 1999; Medway et al., 2011), and respondents who were offered promised incentives were more likely to agree to a urine drug test (Krenzke et al., 2005).

1.4 SUMMARY AND CONCLUSIONS

Survey response rates have been declining in recent years. Because incentives repeatedly have been found to increase survey response rates, they are utilized increasingly in surveys. In particular, prepaid incentives are more effective than promised ones, and monetary incentives are more effective than non-monetary ones. There is some evidence that larger incentives yield greater increases in the response rate; however, there may be diminishing returns from each additional dollar spent. Incentives may be effective at improving the representation of groups that are traditionally hard to reach, such as youth or minorities, as well as people who lack interest in the survey topic or a general interest in participating in research; however, this effect has not been observed across the board. Furthermore, there is mixed evidence as to the utility of incentives for reducing nonresponse bias.

Given the widespread use of incentives, it is important to determine whether incentives affect the level of measurement error in surveys. Fewer studies have looked at measurement effects than at effects on nonresponse, but those that have looked at this issue have generally taken one of three approaches: (1) comparing response distributions, (2) comparing responses to validation data, or (3) comparing respondent effort indicators. These studies typically have concluded that incentives improve the quality of survey data or have no effect on it.

To improve our understanding of incentives' effect on measurement error, we need to move beyond the types of analyses that traditionally have been conducted. For example, comparisons of effort indicators typically have only considered the effect on item nonresponse and responses to open-ended questions; in Chapter 2, I report on the

impact that prepaid cash incentives had on the prevalence of a wider array of satisficing behaviors in a telephone survey. Furthermore, comparisons of response distributions usually have considered all of the survey items as one large group, without any differentiation between types of items; in Chapter 3, I hypothesize that the effect of incentives on responses may vary by item sensitivity and discuss the results of a mail-survey prepaid cash incentive designed to test this possibility. Additionally, few studies have compared survey responses to validation data; in Chapter 3, I also report on the accuracy of responses to three survey items as compared to administrative records. Furthermore, the existing literature rarely examines whether the incentive had a differential impact on measurement error across subgroups of the sample; in these two analytical chapters, I discuss whether the effect of the incentive was restricted to individuals with particular characteristics, such as younger respondents or those with more education. Finally, existing studies typically report on the incentive's effect on each item in isolation; in Chapter 4, I discuss whether prepaid cash incentives affected the relationships between survey responses intended to measure latent characteristics in several recent surveys by testing for measurement invariance between incentive and control group responses.

CHAPTER 2

SATISFICING IN TELEPHONE SURVEYS: DO PREPAID CASH INCENTIVES MAKE A DIFFERENCE?

2.1 INTRODUCTION

Telephone survey response rates have declined considerably in recent years (Brick & Williams, forthcoming; Curtin et al., 2005; Steeh et al., 2001). Incentives are one tool for stemming this decline (Cantor et al., 2008; Curtin et al., 2007; Goetz et al., 1984; Moyer & Brown, 2008; Singer et al., 2000), and, as a result, survey practitioners are often eager to utilize them. However, several studies have found that offering prepaid incentives in telephone surveys can increase the cost per completed interview (Brick et al., 2005; Curtin et al., 2007; Gelman, Stevens, & Chan, 2003; but see Singer et al., 2000). Additional positive outcomes beyond increased response rates may be needed to justify these costs.

If incentives motivate some respondents to participate who otherwise would have declined, they also may influence respondents' behavior *during* the survey interview. Respondents seem more prone to satisfice in telephone surveys than in other modes (Chang & Krosnick, 2009; Hall, 1995; Holbrook et al., 2003), so the potential effect of incentives on respondent effort in telephone surveys is of particular interest. Existing research investigating the effect of incentives on respondent effort in telephone surveys suggests that incentives either result in increased effort or have no effect (e.g., Brehm, 1994; Goetz et al., 1984; Singer et al., 2000). However, these studies are limited in number and examine relatively few indicators of respondent effort. It would be useful to have more evidence about the effect of incentives on respondent effort in telephone surveys.

This chapter describes the methods and results of an experiment using a prepaid cash incentive in a telephone survey. It aims to overcome two limitations of prior research. First, the current study examines the impact of incentives on a wider array of effort indicators than has been considered in the earlier studies. Second, it assesses whether incentives' effect on effort varies according to respondent or item characteristics, whereas most prior research has only discussed their effect on all respondents or items in the aggregate.

2.1.1 Satisficing Theory: Respondent Motivation as a Predictor of Effort

Completing survey interviews can be cognitively taxing for respondents. Though researchers may hope that respondents carefully proceed through all four components of the response process (Tourangeau, Rips, & Rasinski, 2000), cognitive fatigue or lack of interest may lead them to take shortcuts when responding. Instead of responding carefully, respondents may not process survey questions thoroughly and may provide acceptable responses instead of optimal ones (Krosnick, 1991). Satisficing theory proposes a framework for understanding the conditions under which respondents take these cognitive shortcuts (Simon, 1956; Krosnick & Alwin, 1987). According to this theory, the probability that a respondent will satisfice for any given task is a function of three factors – task difficulty, respondent ability to complete the task, and respondent motivation to do so:

$$p(Satisficing) = \frac{a_1(Task\ difficulty)}{a_2(Ability)x\ a_3(Motivation)}$$

Respondents are more likely to satisfice when the task at hand is difficult; however, the greater their ability and motivation, the less likely they are to do so (Krosnick, 1991). Several indicators of satisficing have been proposed, including response order effects,

straight-lining, item nonresponse, and acquiescence (Krosnick, 1991; Krosnick, 1999; Krosnick, Narayan, & Smith, 1996).

Research comparing the quality of responses provided by reluctant respondents with that of the responses provided by those who participate more readily supports the hypothesis that respondents with more motivation may provide higher quality responses than those with less motivation (Cannell & Fowler, 1963; Triplett et al., 1996; Fricker & Tourangeau, 2010; Friedman, Clusen, & Hartzell, 2003; but see Kaminska et al., 2010; Yan et al., 2004). Prepaid cash incentives clearly increase respondents' motivation to take part in surveys; if they also affect respondents' motivation during the interview, they may alter the prevalence of satisficing behaviors.[4]

2.1.2 Theoretical Expectations for Effect of Incentives on Respondent Motivation

Several theories have been offered to explain why prepaid incentives increase response rates. Typically, when researchers have proposed these theories, they have not discussed their implications for the effect of incentives on respondent behavior during the survey interview. Although these theories all are in agreement that incentives should increase response rates, extending them to predict the effect of incentives on respondent motivation and effort during the interview leads to inconsistent predictions about the effects of incentives. The theories suggest four possible effects of incentives on effort: (1) greater effort due to respondents' sense of obligation to the survey researcher, (2) reduced effort due to the influx of uninterested, unmotivated respondents, (3) reduced effort due

[4] Incentives also may affect the likelihood of satisficing by altering the average cognitive ability of the respondent pool. For example, incentives may reduce the proportion of respondents who are older (Dillman, 1996; Miller, 1996; Storms & Loosveldt, 2004); satisficing has been found to be more common among older respondents than among younger ones (Krosnick & Alwin, 1987). Thus, changes in the distribution of cognitive ability need to be taken into consideration when comparing the prevalence of satisficing behaviors among those who have received an incentive and those who have not received one.

to a focus on extrinsic motivations instead of intrinsic ones, and (4) no effect beyond the point of agreeing to participate.

Consider the first of these possibilities – that incentives lead to greater effort due to respondents' sense of obligation to the survey researcher. According to social exchange theory, prepaid incentives may invoke the norm of reciprocity (Gouldner, 1960). When sample members receive an incentive they may feel obligated to return the kindness by responding to the survey (Dillman et al., 2009). A related notion is that incentives create a positive attitude toward the sponsor; some studies find more favorable ratings of the survey sponsor, supporting the hypothesis that prepaid incentives help foster a positive relationship between the sample member and the researcher (e.g., James & Bolstein, 1990). These positive feelings toward the sponsor also could lead respondents to make greater effort than they would have otherwise.

Alternatively, incentives could result in reduced effort due to an influx of uninterested, unmotivated respondents. Incentives may convince certain sample members to respond who are not drawn to other survey features, such as the topic (e.g., Groves et al., 2004). Such respondents may lack sufficient interest and motivation to provide high quality responses.

Additionally, offering incentives could result in reduced effort due to a focus on extrinsic motivations instead of intrinsic ones. The sense of obligation posited by social exchange theory may actually reduce motivation if it causes respondents to focus too heavily on extrinsic reasons for completing the survey. People who are motivated by intrinsic factors, such as interest in the survey topic or enjoyment of sharing one's opinions, may put forth more effort than those who are focused on extrinsic ones, such as

monetary rewards or a sense of obligation. Research on the importance of intrinsic motivation to academic success has supported this assumption (e.g., Bolkan et al., 2011; Fransson, 1977). Similarly, Rush, Phillips, and Panek (1978) report that paid subjects were "striving more for task completion rather than for success" (p. 448). Thus, respondents who see the incentive as their main, or only, motivation for participating may put forth less effort than those who take part for other reasons.

A final possibility is that once sample members have committed to taking part in the survey the incentive has little to no further impact on their behavior. Agreeing to participate in the survey may be sufficient for most sample members to feel they have resolved any cognitive dissonance or met the obligations of the norm of reciprocity. If this is the case, then any further behaviors taken during data collection may not be influenced by the fact that the respondent has received an incentive. Camerer and Hogarth's (1999) review of incentive experiments in laboratory studies found that incentives typically do not affect performance, supporting the view that incentives may not affect behavior during survey interviews.

Additional empirical evidence is needed to help determine which of these expectations is the most accurate, and whether our expectations should vary according to survey design features or respondent characteristics.

2.1.3 Existing Studies Investigating the Impact of Incentives on Respondent Effort

Most studies investigating the impact of incentives on respondent effort focus on item nonresponse. Several have observed a reduction in item nonresponse when incentives are utilized (Godwin, 1979; Goetz et al., 1984; Goldenberg et al., 2009; James & Bolstein, 1990; McDaniel & Rao, 1980). However, many others have concluded that

incentives do not have a significant impact on the rate of item nonresponse (e.g., Berk et al., 1987; Berlin et al., 1992; Cantor et al., 2008; Curtin et al., 2007; Davern et al., 2003; Furse & Stewart, 1982; Peck & Dresch, 1981; Shettle & Mooney, 1999). It is unclear what design characteristics lead to these divergent results; the studies in both groups have used prepaid and promised incentives, have been conducted in a variety of modes, and have asked respondents about a wide range of topics. None of the incentive experiments I found resulted in a significant overall increase in item nonresponse.

The other effort-related outcome that frequently has been analyzed in incentive studies is the quality of open-ended responses, generally operationalized as the number of words or number of ideas included in the response. Several studies have concluded that incentives lead to an improvement in quality (Brehm, 1994; Goetz et al., 1984; Willimack et al., 1995), although multiple studies that have offered more than one value of monetary incentive have found that only the larger amount has resulted in improved response quality (Godwin, 1979; James & Bolstein, 1990; Tzamourani & Lynn, 1999). I only came across one study where the incentive led to a significant reduction in response quality (Hansen, 1980).

Only a few incentive studies have considered effort indicators beyond item nonresponse and responses to open-ended items; again, the studies that have done so have generally found that incentives either improve effort or have no effect. For example, the number of events recorded in a diary was increased when an incentive was provided (Lynn & Sturgis, 1997). However, a prepaid voucher did not have a significant effect on the number of imputations or edits required for 40 items in the Survey of Income and Program Participation (Davern et al., 2003). In two other recent surveys, prepaid cash

incentives did not affect the number of responses selected for a check-all-that-apply item, the proportion of respondents who provided at least one pair of inconsistent responses, or the proportion of respondents who provided round numerical responses (Medway et al., 2011). Finally, a few studies have found that respondents who have received incentives have been more willing to submit to requests that imply potential additional burden; for example, respondents who received prepaid incentives were more likely to provide additional contact information (Shettle & Mooney, 1999; Medway et al., 2011), and respondents who were offered promised incentives were more likely to agree to a urine drug test (Krenzke et al., 2005).

These studies provide information about how incentives affected effort across all items for all respondents; however, it is possible that the effect of the incentive is restricted to certain subgroups of respondents or particular types of items. Only a few studies have examined these possibilities. For example, Singer et al. (2000) found that incentives in the Survey of Consumers led to a significant reduction in item nonresponse within two particular subgroups – older respondents and non-Whites; however, this result was not replicated in a similar experiment in a subsequent administration (Curtin et al., 2007). Similarly, McDaniel and Rao (1980) found that a prepaid incentive significantly reduced item nonresponse for both open-ended and closed items, while Hansen (1980) found that the effect of incentives was limited to open-ended items. Finally, Tzamourani and Lynn (1999) concluded that providing a prepaid monetary incentive reduced item nonresponse for non-sensitive questions but increased it for sensitive ones, while Krenzke and colleagues (2005) found that offering a promised cash incentive did not have a significant impact on the proportion of items respondents skipped, regardless of item

sensitivity. In this same study, the researchers hypothesized that the effect of the incentive would be greater in the final section of the interview, when respondents were tired, but this prediction was not supported by the data.

2.1.4 Extending the Literature

As this review shows, most of the existing incentive experiments have focused on item nonresponse as the primary indicator of respondent effort. Item nonresponse rates are an attractive indicator of data quality in the sense that they are easily calculated and compared across studies. Furthermore, reducing item nonresponse is desirable because it decreases the amount of imputation that must be done. However, while the level of item nonresponse is a widely used indicator of respondent effort in survey research, it only captures the absolute minimum amount of information about respondent effort; it tells researchers that the respondent took the time to provide an answer, but it tells them nothing about the actual quality of that response. Several other indicators of effort that hold respondents to a higher standard, such as response order effects and non-differentiation, have been utilized in other areas of survey research but have not been applied to incentives research. Measuring the impact of incentives on such indicators would improve researchers' knowledge of the degree to which incentives influence respondent effort. The current study includes measures of twelve indicators of respondent effort; the operationalization of each indicator is discussed at greater length in the Methods section, but most of them are derived from the literature on survey satisficing.

Furthermore, the current study examines the possibility that the effect of incentives varies according to respondent or item characteristics. Significant effects at the subgroup level may be masked at the aggregate level. I hypothesize that an incentive will

increase effort among respondents who recall receiving it but not among other respondents. I also hypothesize that incentives will have a greater impact on respondents with higher levels of cognitive ability and lower levels of conscientiousness. Indicators of these characteristics are included in the current study and are discussed in further detail in the Methods section of this chapter. Finally, I hypothesize that the effect of the incentive will be greater for two types of items. First, due to cognitive fatigue, the incentive will have a greater impact on effort in the second half of the interview than it will in the first half. Second, because answers to attitude items are more affected by context than answers to factual or behavioral questions, the incentive will have a greater effect on responses to the former than on responses to the latter.

2.2 RESEARCH METHODS

2.2.1 Sampling Frame and Experimental Conditions

The data for this study come from the 2011 JPSM Practicum survey. As part of this study, a telephone survey was conducted in the summer of 2011 by Princeton Survey Research Associates International (PSRAI). The target population was non-institutionalized persons age 18 and older living in the continental United States. Survey Sampling International (SSI) provided a sample of 9,500 individuals. SSI creates its directory-listed files by merging directory-listed residential telephone numbers with a variety of secondary sources, such as birth records, voter registrations, and motor vehicle registrations. SSI selected the sample for this study so that the number of records selected for each state and county was in line with Census population distributions.

A listed sample was chosen because it included a name, address, and phone number for each case. An address and a phone number were needed to send each sample member an advance letter and interview him or her on the telephone. Having full name information for each sample member increased the power of the incentive treatment; the advance letter was addressed to the specifically-named individual listed in the sample file, and only this individual was eligible to complete the telephone interview. This increased the likelihood that the individual who completed the interview also was the household member who had read the advance letter and received the cash incentive.

7,200 sample members were randomly selected from the initial list of 9,500 cases received from SSI. Just over one percent of the 9,500 cases did not include a first name; because of the difficulty of requesting to speak with a person for whom we did not have a first name, these cases were dropped from the file prior to selecting the sample. SSI

indicated that an additional one percent of the 9,500 cases were ported numbers; to avoid inadvertently autodialing wireless numbers, these cases also were removed from the file before the sample was selected.

All 7,200 sample members were sent an advance letter. The letters were released in two replicates. The first batch was sent to 3,400 sample members on July 14-15, 2011, and the second was sent to 3,800 sample members on July 28-29, 2011. As part of this experiment, 40% of the sample members were randomly assigned to receive a $5 prepaid incentive with the advance letter. The other 60% of the sample received an advance letter without an incentive. Both replicates were included in the experiment. The exact wording of the advance letter is included in Appendix A.

Interviews were conducted between July 18 and August 17, 2011. PSRAI made up to six attempts to reach sample members. Nine hundred interviews were completed. The median interview length was 19.5 minutes. The survey covered several topics, including health, employment, and current social issues.

2.2.2 Indicators of Respondent Effort

The survey questionnaire included measures of several indicators of respondent effort. First, it included measures of the two indicators most commonly studied in prior incentive experiments: (1) item nonresponse and (2) responses to open-ended items. The survey also included measures of other traditional satisficing indicators originally proposed by Krosnick and colleagues (Krosnick, 1991; Krosnick, 1999; Krosnick et al., 1996): (3) straight-lining and non-differentiation, (4) acquiescence, and (5) response order effects. Finally, the survey included indicators that survey researchers have used to determine respondents' level of effort in other contexts: (6) lack of attention to important

exclusions, (7) use of round or prototypical values for numerical responses, (8) use of estimation strategies to answer questions requiring numerical responses, (9) underreporting to filter items, (10) interview length, and (11) accuracy of survey reports as compared to frame data. After the call was completed, (12) the interviewers provided observations about each respondent's level of effort during the interview.

With the exception of accuracy of survey reports as compared to frame data, these are indirect indicators of measurement error. Using round numbers for numerical responses and providing brief responses to open-ended items does not prove that an individual's responses are prone to extensive measurement error, but it does imply that he or she may be making less effort; by extension such individuals may also provide less accurate responses. In the section below, I provide more information about each of these indicators, including how they were measured in the questionnaire and how I analyzed the data. Exact question wordings can be found in Appendix A, while information about which items were included in each indicator is located in Appendix B.

Item nonresponse. When respondents feel that a survey item is too cognitively burdensome, they may decline to provide an answer instead of formulating a response. To determine the effect of the incentive on item nonresponse in the current study, I calculated the proportion of the items that each respondent declined to answer.[5] If respondents who received an incentive made greater effort than those who did not receive one, they should have skipped a smaller proportion of the items.

[5] The questionnaire included several experiments and skip patterns. As a result, most of the respondents were not asked all of the survey questions. Because of this, I often report outcomes as proportions – the number of times the respondent did a certain behavior divided by the number times they had the opportunity to do so.

Responses to open-ended items. Open-ended items, in which respondents are asked to come up with their own response instead of choosing one from a pre-selected list of options, can be especially burdensome. The current study included three such items. To determine the effect of the incentive on responses to open-ended items, I calculated the number of words that each respondent provided in response to each item. For one of the items I also was able to calculate and compare the number of ideas that respondents provided as part of their answer. Finally, I determined the proportion of the respondents who skipped each of these items. If respondents who received an incentive made greater effort than those who did not receive one, they should have provided more words and ideas on average and fewer of them should have skipped these items.

Straight-lining or non-differentiation. Respondents may be presented with a series of items that have the same response options. Straight-lining occurs when respondents select the same response for all of the items in such series, and non-differentiation occurs when respondents select very similar responses for all of the items (Krosnick, 1991). To measure the effect of the incentive on these behaviors, I included three multi-item batteries in the questionnaire. Respondents were considered to have straight-lined if they provided the same response to all of the items in a particular battery. To determine the degree of differentiation in responses, I calculated the standard deviation of each respondent's answers to each battery. If respondents in the incentive group made greater effort, a significantly smaller proportion of them should have straight-lined, and there should be significantly more differentiation in their responses to batteries of items.

Acquiescence bias. Respondents are often asked questions that require either an affirmative or negative response. These statements typically ask about things that are reasonable for respondents to believe or do. Individuals who are not processing these questions deeply may find it easier to find reasons to provide an affirmative response than to provide a negative one (Krosnick, 1991). The current study included 29 such items. I calculated the number of affirmative responses each respondent gave to these items. If respondents who received an incentive made greater effort than those who did not, they should have provided a smaller number of affirmative responses to these items.

A drawback to this approach is that differences between the incentive and control group in the average number of affirmative responses could be due to true differences in the prevalence of the attitudes or behaviors asked about in these items. As a result, I included wording experiments for four of the items in which one half of the respondents was asked the original version of the question and the other half was read a question asking about the opposite attitude. For example, one half of the respondents was asked whether they agree or disagree that, "Increasing government involvement in healthcare will improve the quality of care," while the other half was asked if this will *hurt* the quality of care. If respondents in the incentive group made greater effort, they should be less likely than control group respondents to agree with both versions of the item.

Response order effects. Respondents who are satisficing may also demonstrate response order effects. In aural modes such as the telephone, both recency effects and primacy effects may occur (Krosnick & Alwin, 1987). To assess the effect of the incentive on response order effects, I examined the 27 items in the questionnaire that had at least four response options. I determined the proportion of these items for which for

each respondent did each of the following: selected one of the first two options, selected the first option, selected one of the last two options, or selected the final option. If respondents who have received an incentive made greater effort than those who did not, then they should have selected the first or last options less frequently.

Again, the drawback to this approach is that differences in the observed prevalence of primacy or recency effects between the incentive and control group could potentially be due to true differences in the prevalence of the attitudes or behaviors asked about in these items. Therefore, four experiments were included in the questionnaire that varied the order of the response options; one half of the respondents received the initial order and the other half received the reverse order. If respondents who received an incentive made greater effort than those who did not receive one, then they should have selected the first or last options less often, regardless of presentation order.

Lack of attention to question wording. If respondents are taking cognitive shortcuts they may not listen very carefully to the reading of survey items and may miss important details such as instructions to include or exclude certain categories from consideration in their response. In the current study, four experiments were conducted to determine the effect of the incentive on attention to question wording. For each item, respondents were randomly assigned to either the standard wording or the exclusion wording. For example, respondents in the standard condition were asked, "In the past seven days, how many servings of vegetables did you eat?" The exclusion condition respondents also were told, "Please do not include carrots, beans, or lettuce". For each item, I determined the mean response provided in the standard and exclusion conditions and calculated the difference between these two means. If respondents who received an

incentive made greater effort than those who did not receive one, the difference between the standard and exclusion groups should be significantly greater among respondents in the incentive group.

Use of round or prototypical values for numerical responses. Respondents often are asked to provide numerical responses to open-ended items. Respondents who are taking cognitive shortcuts may report round or prototypical numbers (such as multiples of 5 or 10) instead of making the extra effort to report an exact numerical response. The current study included 15 items that requested numerical responses. For all of the items, respondents who provided a value that was a multiple of five were considered to have given a round response. In addition, some items asked how often respondents did something in a week or a year; for these items, multiples of seven or twelve also were considered round responses. I determined the proportion of items for which each respondent provided a round response. If respondents who received an incentive made greater effort than those who did not receive one, then they should have provided round responses for a smaller proportion of the items.

Estimation strategy for numerical responses. When respondents are asked to provide a number that indicates how often they perform a certain behavior, there are several approaches they can use to come up with an answer that require varying levels of effort (Tourangeau et al., 2000). Respondents who are trying to conserve cognitive energy may estimate their response instead of making the effort to recall each particular episode (Conrad, Brown, & Cashman, 1998). To assess the effect of the incentive on recall strategy, I asked respondents to indicate how they came up with their answer to one such question, in which they were asked how many times they had seen a doctor or other

health professional in 2010. Respondents who said they had seen a doctor at least once were then asked, "Which of the following describes how you came up with your answer: did you estimate based on a general impression; did you think about types of visits; did you think about how often you usually go to the doctor; or did you think about each individual visit?" Respondents could select multiple responses. Any response other than thinking about each individual visit suggested that respondents had estimated their response. If respondents who received an incentive made greater effort than those who did not receive one, then fewer of them should have estimated their response.

Underreporting to filter items. Surveys often include groups of questions that consist of an initial filter question and a set of follow-up questions for those whose answers to the filter question indicate that the follow-up items apply to them. Interleafed formats are a relatively popular way to organize such questions in survey questionnaires. In this format, the follow-up items come immediately after the relevant filter question. With this format, however, respondents may learn that negative answers to the filter questions help them end the interview more quickly. Respondents who are trying to reduce their cognitive burden may then provide false negative responses to these filter items (Jensen, Watanabe, & Richter, 1999; Kreuter, McCulloch, Presser, & Tourangeau, 2011; Lucas et al., 1999).

To determine the effect of the incentive on reports to filter items, I included a section in the questionnaire that consisted of six filter questions (and their follow-ups) in an interleafed format. Respondents were asked whether they had ever been diagnosed with six medical conditions, such as hypertension or diabetes. The order of these questions was randomized. If respondents replied affirmatively they were asked at what

age they had been diagnosed. This question may be burdensome enough that some respondents may have wanted to avoid answering it. As a result, they may have started responding negatively to the diagnosis questions that were presented later in the list. If respondents who received an incentive made greater effort than those who did not receive one, then presentation order should have a significantly smaller impact on their likelihood of providing an affirmative response.

Interview length. Respondents who are fatigued may try to rush through the survey interview. Such respondents may listen to the survey questions less carefully and make less effort when formulating a response. For example, Malhotra (2008) found that overall completion time is a significant negative predictor of primacy effects for some subgroups of respondents. To determine the effect of the incentive on interview length, the start and end time of each interview was recorded. If respondents who received an incentive made greater effort than those who did not receive one, then their mean response time should be longer than that of the control group.

Accuracy of survey reports as compared to frame data. All of the other indicators of respondent effort are indirect indicators of measurement error. In comparing respondent reports to data from the frame, we have a more direct indicator of measurement error – whether the self-report matches the frame, and, if it does not, the size of the discrepancy between the two pieces of information. To determine whether the incentive affected the accuracy of self-reports, I included one question that asked respondents for information that was also available on the frame – how many years they have lived in their current residence. Although there is likely some error in the frame data, we can assume that these errors are equally present for the incentive and control

groups due to the random assignment of sample members to experimental conditions. I compared survey reports to the frame data to determine the proportion of respondents whose answers matched the frame exactly. I also calculated the mean discrepancy between self-reports and frame data in the two groups. If respondents who received an incentive made greater effort than those who did not receive one, there should be a smaller number of discrepancies in the reports of incentive respondents and the mean discrepancy in reports should be smaller.

Interviewer reports of effort. Through their interaction with the respondent throughout the interview, interviewers can formulate an impression of the amount of effort made by respondents. These impressions can incorporate effort indicators that are not captured by survey responses – such as respondents' decision to consult records during the interview or indications that the respondent is distracted by other things happening in the room. To determine whether the incentive affected interviewers' impression of respondent effort, the interviewers rated each respondent's effort after the call was completed; they recorded the extent to which "the respondent answered the survey questions to the best of their ability," on a five point scale from "not at all" to "very often". If respondents who received an incentive made greater effort than those who did not receive one, interviewers should have rated their effort as being greater than that that of the control respondents.

2.2.3 Respondent Characteristics

Prior studies investigating the impact of incentives on data quality have found weak effects or no effects at all. This absence of findings may be due to the fact that existing studies have considered the effect of incentives on all respondents in the

aggregate. The effect of incentives on effort may be limited to certain subgroups of respondents. Analyzing all respondents at once may mask the effect the incentive has on these subgroups. The current study's questionnaire included measures of several respondent characteristics hypothesized to impact whether an incentive would affect the level of effort put forth by respondents.

Cognitive ability. I hypothesized that receiving an incentive would have little effect on the prevalence of satisficing among respondents with lower cognitive ability – because of their lower cognitive ability, these respondents cannot help satisficing; in contrast, the incentive should have a greater impact on the behavior of respondents who have fewer cognitive constraints. This study included three indicators of cognitive ability. The first was age; respondents over the age of 65 were considered to have lower cognitive ability. The second was education; respondents who reported having a high school diploma or less were considered to have lower cognitive ability. The final indicator was interviewer reports of respondent difficulty in answering the survey questions. After the interview was completed, the interviewers were asked to report how often the respondent had trouble understanding the survey questions on a five-point scale from "not at all" to "very often"; respondents who had trouble understanding the questions somewhat often, pretty, often, or very often were considered to have lower cognitive ability.

Conscientiousness. I hypothesized that the more internally motivated respondents were, the smaller the effect of the incentive would be on their effort during the interview. In the current study, I used self-reported conscientiousness as a proxy for intrinsic motivation. Conscientiousness is one of the five personality traits on which individuals

are said to vary on the "Big 5" personality inventory; people who rate highly on conscientiousness tend to be thorough, self-disciplined, and pay attention to details (Costa & McCrae, 1992). Respondents who are conscientious should be motivated to complete the questionnaire at a high level of quality regardless of whether they have received an incentive; conversely, the motivation and dedication of respondents who are not very conscientious may be affected by whether or not they have received an incentive. The questionnaire included 10 items intended to measure conscientiousness; these items were adapted from a measure developed by Buchanan, Johnson, and Goldberg (2005). Respondents were asked to what extent they agreed or disagreed that each item described them on a five-point scale from "strongly disagree" to "strongly agree". Example items include, "I am always prepared," and, "I do just enough work to get by". I calculated the mean score that respondents gave across the items, with higher means indicating greater levels of conscientiousness.

Incentive recall. Finally, I hypothesized that the incentive would have little impact on the behavior of respondents who did not recall receiving it. Although the advance letter was addressed to a specific individual and this same person was asked to complete the telephone interview, it remained possible that some of the respondents in the incentive group were not aware of the incentive. To measure incentive recall, respondents first were asked, "A letter describing this study may have been sent to your home recently. Do you remember seeing the letter?" Respondents who recalled the letter were asked an open-ended follow-up, "Do you happen to remember if there was anything else in the envelope with the letter?" If respondents said "yes", interviewers were instructed to probe to determine what the respondent remembered being in the envelope.

PSRAI also kept track of advance letters that were returned because they were undeliverable; the respondents to whom these letters were sent were considered to be unaware of the advance letter and incentive.

2.3 RESULTS

This section presents the results of the incentive experiment described above. First, I discuss the effect of the incentive on the response rate, cost per complete, and sample composition. Then, I review the effect of the incentive on several indicators of respondent effort. Finally, I discuss whether the effect differed depending on respondent characteristics.

2.3.1 Outcome Rates

Table 2.1 provides an overview of the outcomes of the data collection effort. The overall response rate was 15.7% (AAPOR RR1). The response rate was significantly higher in the incentive condition (22.8%) than it was in the control condition (10.9%). The cooperation rate also was significantly higher in the incentive condition than in the control condition (39.2% and 20.0%, respectively), while the refusal rate was significantly lower (32.4% and 41.0%, respectively). Although not anticipated, the contact rate also was significantly higher in the incentive condition than it was in the control condition (58.1% and 54.7%, respectively). This may have been the result of significantly increased advance letter recall in the incentive group (86.5% and 66.0%, respectively); the advance letter mentioned the name of the survey research firm that would be calling, and improved recall of the letter may have led more of the incentive sample members to recognize the name when it appeared on caller ID.

Table 2.1. Outcome Rates, by Incentive Condition

	Total	Incentive	Control	$\chi^2(1)$
Response Rate (RR1)	15.7%	22.8%	10.9%	145.58***
Refusal Rate (REF1)	37.6%	32.4%	41.0%	44.43***
Contact Rate (CON1)	56.1%	58.1%	54.7%	6.42*
Sample size	(7,199)[A]	(2,880)	(4,319)	
Cooperation Rate (COOP1)	28.0%	39.2%	20.0%	142.61***
Sample size	(3,216)	(1,337)	(1,879)	
Remembered advance letter	77.9%	86.5%	66.0%	53.38***
Sample size	(900)	(524)	(376)	

[A] Due to an error with the call scheduling software, one sample case was never called.
*p<0.05, **p<0.01, *** p<0.001

2.3.2 Cost per Complete

Because the same interviewers worked both the incentive and control cases simultaneously, it is impossible to determine the exact cost per complete by incentive condition. However, I was able to estimate the cost per complete using information about the number of calls made to the incentive and control conditions. 8,037 calls were made to obtain 376 control interviews – equivalent to 21.4 calls per complete. Meanwhile, 5,290 calls were made to obtain 524 incentive interviews – equivalent to 10.1 calls per complete. By subtracting the cost of the incentives and advance letters from the total project cost, I determined the total cost of making these phone calls. Because 60% of the calls were made to control group cases, I allocated 60% of the total calling costs to the control group and 40% to the incentive group. Next, I distributed the incentive costs (100% to incentive group) and advance letter costs (40% to incentive group, 60% to control group). I summed these three costs for each experimental condition and divided this sum by the number of completes achieved in that condition. As shown in Table 2.2, the cost per complete was greater in the control condition ($63.76) than it was in the incentive condition ($57.68). Similar to Singer et al. (2000) but in contrast to other recent

studies (Brick et al., 2005; Curtin et al., 2007; Gelman et al., 2003), incentives appear to have been the more cost-effective approach for this survey.

Table 2.2. Costs (Dollars), by Incentive Condition

	Total	**Incentive**	**Control**
Advance letters	9,144.00	3,657.60	5,486.40
Incentives	14,400.00	14,400.00	0.00
Phone calls	30,656.00	12,168.55	18,487.45
Total	54,200.00	30,226.15	23,973.85
Cost per complete	60.22	57.68	63.76

2.3.3 Sample Composition

As compared to the March 2010 Current Population Survey (CPS), the Practicum survey underrepresented females, people under the age of 65, non-Whites, and those with lower levels of education (Figure 2.1). These discrepancies were due in part to the use of a listed sample. Such samples do not include cell telephone numbers, so they are prone to underrepresent subgroups that are likely to be cell-only, such as younger people, non-Whites, those living in urban areas, and those with less education (Blumberg & Luke, 2007; Keeter, Kennedy, Clark, Tompson, & Mokrzycki, 2007). The underrepresentation of females was likely due to the decision to contact the specifically-named household member listed on the frame; this was done in order to increase the likelihood that the person who received the incentive would also complete the survey. The person listed on the sampling frame was the head of the household, and, among heterosexual couples living together, the head of household is more likely to be male than to be female (Kleinjans, 2008).

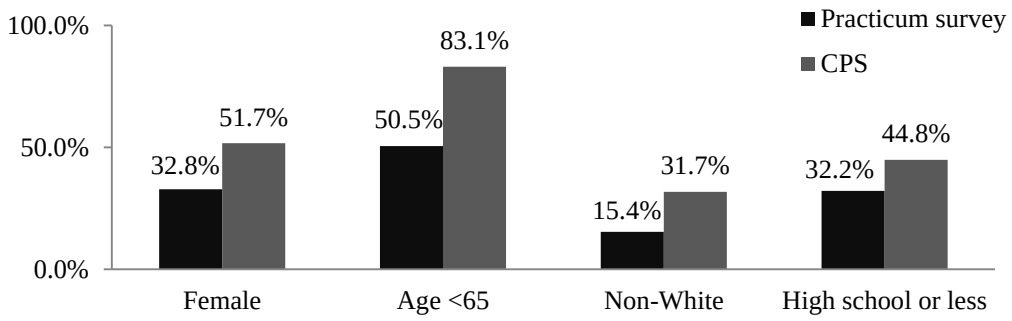

Figure 2.1. Demographic Characteristics, by Survey

Although the respondents to this survey may not be representative of the entire United States population, they more accurately represent the types of people to whom prepaid incentives can be sent in most random-digit-dial surveys. Researchers hoping to use prepaid cash incentives in such surveys can only send them to sample cases for which they can link the phone number to an address. These are typically older, non-Whites of higher socioeconomic status, similar to the respondents to the Practicum survey (Link & Mokdad, 2005). Thus, the results of this study offer a good representation of the effect that prepaid incentives have on satisficing behavior *among those sample members to whom they can realistically be sent.* As shown in Table 2.3, there were not any significant differences in the composition of the incentive and control groups in terms of gender, age, education, race/ethnicity or region.

Table 2.3. Sample Composition, by Incentive Condition

	Incentive *(Percent)*	**Control** *(Percent)*	**Chi-Square**
Gender (n=900)			$\chi^2(1)=1.97$
Male	69.1	64.6	
Female	30.9	35.4	
Age (n=871)			$\chi^2(6)=4.25$
18-34	3.8	3.6	
35-44	8.5	6.3	
45-54	16.4	14.6	
55-64	23.1	24.5	
65-74	24.1	22.0	
75-84	18.0	21.4	
85+	6.3	7.7	
Race/ethnicity (n=885)			$\chi^2(3)=4.11$
White, non-Hispanic	86.2	82.4	
Black, non-Hispanic	4.3	7.3	
Hispanic	3.5	3.5	
Other race, non-Hispanic	6.0	6.8	
Education (n=895)			$\chi^2(5)=5.20$
Less than high school	5.2	4.0	
High school	28.0	26.8	
Some college, no degree	19.5	22.5	
Associate's degree	8.6	9.4	
Bachelor's degree	21.3	22.3	
Graduate degree	17.4	15.0	
Region (n=898)			$\chi^2(3)=4.06$
Northeast	14.3	17.1	
Midwest	26.2	21.3	
South	34.4	37.9	
West	25.0	23.7	

Note: Due to rounding, proportions may not sum to 100%.

2.3.4 Effect of Incentive on Respondent Effort Indicators

Item nonresponse. Just over half of the respondents did not answer at least one survey question (57.0%). As shown in Table 2.4, respondents in the incentive group were significantly less likely than those in the control group to skip at least one item (54.0% and 61.1%, respectively). Respondents in the incentive group also skipped a significantly

smaller proportion of the items on average (2.3% and 2.9%, respectively).[6] These results remained significant after controlling for whether or not the respondent was age 65 or older. However, among respondents who skipped at least one item, there was not a significant difference in the proportion of the items that were skipped (4.2% and 4.7%, respectively). Repeating these analyses with the nonresponse proportions capped at the 99th percentile did not change the direction or significance of the results.[7]

Respondents might skip an item for several reasons. If the reduction in item nonresponse among incentive respondents is due to increased effort, we should expect the reduction in nonresponse to be greater for "don't know" responses than for "refuse" responses. As shown in Table 2.4, incentive group respondents were significantly less likely than control group respondents to provide at least one "don't know" response (47.8% and 56.1%, respectively), and they provided a "don't know" response to marginally significantly smaller proportion of the items (1.5% and 1.9%, respectively; p=0.06). Conversely, the incentive did not have a significant effect on the proportion of respondents who refused at least one item or the mean proportion of items that elicited a refusal. Controlling for whether or not the respondent was age 65 or older did not change the results of these analyses. Although the incentive appeared to have a greater effect on "don't know" responses than it did on "refuse" responses, a repeated-measures analysis indicated that the effect of the incentive on the proportion of "don't know" responses was not *significantly* greater than the effect on the proportion of refusals.

[6] Respondents were asked between 73 and 82 questions, depending on which items they skipped automatically; as a result, I focus on the questions they chose to skip as proportion of the questions they actually received.

[7] Where applicable, all subsequent analyses were repeated with the largest values trimmed at the 99th percentile. I only report on these analyses if they changed the direction or significance of the results.

Table 2.4. Item Nonresponse, by Incentive Condition

	Incentive *(Percent)*	**Control** *(Percent)*	**Significance**
Respondent skipped at least one item	54.0	61.1	$\chi(1) = 4.58$*
At least one "don't know" response	47.8	56.1	$\chi(1) = 7.86$**
At least one "refuse" response	27.5	32.2	$\chi(1) = 2.33$
Sample size	(524)	(376)	
Mean proportion items skipped	2.3	2.9	t = -2.01*
First half	3.0	3.4	t = -1.42
Second half	1.7	2.5	t = -2.16*
Mean proportion "don't know"	1.5	1.9	t = -1.91
First half	2.0	2.4	t = -1.37
Second half	1.1	1.6	t = -2.00*
Mean proportion "refuse"	0.8	1.0	t = -1.21
First half	1.0	1.1	t = -0.82
Second half	0.6	0.9	t = -1.16
Sample size	(524)	(376)	
Mean proportion items skipped (if skipped at least 1 item)	4.2	4.7	t = -1.11
Mean proportion "don't know"	2.8	3.2	t = -1.05
Mean proportion "refuse"	1.5	1.6	t = -0.56
Sample size	(283)	(230)	

*p<0.05, **p<0.01

Finally, I hypothesized that the incentive might have a greater effect on respondent effort in the second half of the interview, as respondents grew more fatigued. As shown in Table 2.4, the incentive led to significantly fewer skipped items and "don't know" responses in the second half of the interview but did not have a significant impact in the first half of the interview;[8] however, a repeated-measures analysis indicated that the effect of the incentive on the proportion of items skipped was not *significantly* greater in the second half than in the first half.

[8] This analysis was repeated with nonresponse proportions trimmed at the 99th percentile; with this change, the difference in the mean proportion of "don't know" responses provided by incentive and control group respondents in the second half was only marginally significant (1.1% and 1.4%, respectively; t=-1.91; p=0.06), and all other results remained the same.

Quality of responses to open-ended items. Open-ended items are more cognitively burdensome for respondents than are closed items. One way to reduce the effort required for answering an open-ended item would be to not provide any answer at all. As shown in Figure 2.2, there was not a significant difference in the proportion of respondents in the incentive and control groups that skipped each of the open-ended items.

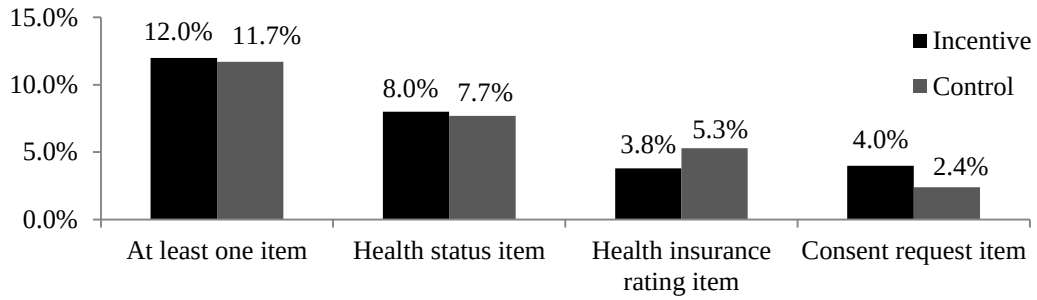

Figure 2.2. Item Nonresponse for Open-Ended Items, by Incentive Condition

The number of words provided in the response is another indication of the effort the respondent put into answering the question. As shown in Table 2.5, there was not a significant difference between the mean number of words provided in the incentive and control groups for any of the three items. There also was not a significant difference in the mean number of words provided across all three items in the incentive and control groups (13.0 and 13.8 words, respectively).

Table 2.5. Mean Number of Words Provided, by Incentive Condition

	Incentive	**Control**	***t-value***
Across all three items	13.0 (n=524)	13.8 (n=376)	-1.50
Health status item	9.6 (n=523)	9.7 (n=373)	-0.13
Health insurance rating item	11.6 (n=495)	11.4 (n=348)	0.33
Consent request item	9.9 (n=515)	10.6 (n=372)	-1.28

Note: For all calculations presented in this table, respondents who said "refuse" or "don't know" were considered to have provided zero words. Respondents who were not asked a particular question were excluded from the associated analysis.

Finally, respondent effort was estimated based on the number of unique ideas that the respondent mentioned. For the consent request item, the open-ended responses were coded into 21 different categories. Respondents expressed up to four ideas each, but the majority of them expressed only one idea (Table 2.6). There was not a significant difference in the distribution of responses by the number of ideas they included, nor in the mean number of ideas provided in the incentive and control groups (1.1 and 1.2, respectively; p=0.30). Overall, these analyses suggest that the incentive had no effect on the quality of the responses provided to the open-ended items.

Table 2.6. Number of Unique Ideas Provided for Consent Explanation, by Incentive Condition

	Incentive *(Percent)*	**Control** *(Percent)*
Zero ideas	4.0	2.4
One idea	80.3	79.3
Two ideas	14.5	16.8
Three or more ideas	1.2	1.6
Sample size	(503)	(367)

Note 1: Respondents who said "don't know" or "refused" are considered to have provided zero ideas. Respondents who were not asked the question were excluded from the analysis.
Note 2: $\chi^2(4)$=3.92; p=0.42

Straight-lining / non-differentiation. Respondents were presented with three series of items. The first included five items asking about confidence in various institutions in American society on a 10-point scale, the second included 10 items asking about patriotism on a four-point scale, and the final series included 10 items asking about the respondent's conscientiousness on a five-point scale. Straight-lining – providing the same response for all of the items in the series – was rare; only 3% of respondents exhibited this behavior. Among those who did straight-line, the vast majority did so for only one of the three batteries; only one respondent straight-lined for more than one battery. The incentive did not significantly affect the prevalence of straight-lining behavior (Table 2.7).

Table 2.7. Straight-lining and Non-differentiation, by Incentive Condition

	Incentive	**Control**	**Significance**
Straight-lining	*Percent Straight-lining*	*Percent Straight-lining*	*Chi-Square*
At least one battery	2.3	3.7	2.84
Battery 1: Confidence in institutions	1.0	1.6	0.75
Battery 2: Patriotism	1.3	2.1	0.84
Battery 3: Conscientiousness	0.2	0.0	0.72
Non-differentiation	*Mean Standard Deviation*	*Mean Standard Deviation*	*t-value*
Across all three batteries	1.46	1.45	0.32
Battery 1: Confidence in institutions	1.98	1.95	0.59
Battery 2: Patriotism	0.74	0.72	1.34
Battery 3: Conscientiousness	1.65	1.68	-1.08
Sample size	(524)	(376)	

Note: All chi-square tests based on 1 degree of freedom

Straight-lining is an extreme form of satisficing; a more subtle cognitive shortcut is for respondents to consider only a subset of the response options. This is known as non-differentiation and is measured by looking at the variation in the responses that each respondent provides for the items in a given battery. To determine the effect of the

incentive on non-differentiation for each series, I computed the standard deviation of each respondent's answers to each series. The incentive did not have a significant impact on the differentiation in respondents' answers for any of the three batteries individually or on average across all three batteries (Table 2.7). Overall, these results suggest that the incentive had no impact on the variability of the responses provided to batteries of items.

Acquiescence. Respondents to this survey were asked 29 questions to which they were expected to provide either an affirmative or negative response. On average they responded affirmatively to 12 of them. As shown in Table 2.8, there was not a significant difference in the average number of affirmative responses provided by respondents in the incentive and control groups (11.9 and 12.0, respectively). I hypothesized that the effect of the incentive might be limited to attitudinal questions since these are more context-dependent than behavioral items; however, there was not a significant difference in the average number of affirmative responses to either type of item. I also hypothesized that the incentive might have a greater effect in the second half of the interview than in the first half, but there was not a significant difference in the number of affirmative responses provided in either half of the interview.

I also compared the proportion of respondents in each group that provided an affirmative response to each item. Incentive group respondents were less likely to provide an affirmative response for three-fourths of the behavioral items (8 of 11) and three-fifths of the attitudinal items (11 of 18). However, these differences generally were very small and none of them were statistically significant at the 0.05 level.[9]

[9] Respondents in the incentive group were significantly more likely than those in the control group to agree to the consent request at the $p<0.10$ level (32% and 29%, respectively; $\chi 2(1)=2.82$; $p=0.09$). Looking at the responses for the two versions of the consent request separately showed that this difference was driven by a significantly greater proportion of the incentive respondents giving consent to access their health records

Table 2.8. Mean Number of Affirmative Responses, by Incentive Condition

	Incentive	**Control**	***t*-value**
Overall	11.9	12.0	-0.30
Attitudinal items	8.5	8.5	0.16
Behavioral items	3.4	3.5	-0.29
First half	5.3	5.4	-0.42
Second half	6.6	6.6	0.62
Sample size	(524)	(376)	

The drawback to the analyses discussed thus far is that they make it difficult to isolate acquiescence from true differences in the attitudes and behaviors of the incentive and control groups. As a result, four attitudinal items were included in a wording experiment aimed at overcoming this shortcoming. There were two versions of each item that expressed opposite opinions (e.g., "Increasing government involvement in health care will [improve/hurt] the quality of care."). One half of the respondents in the control and incentive groups received each wording. It is unlikely that the incentive group would have both more people who support increased government involvement in healthcare than the control group *and* more who oppose it.

If a significantly smaller proportion of incentive respondents agree with both question versions, this suggests that there was less acquiescence in the incentive group than in the control group. As a test of this, I estimated separate logistic regressions for each of the versions for all four items included in the experiment, where the dependent variable was whether or not the respondent provided an affirmative response to the item and the independent variable was incentive condition. Table 2.9 shows the odds ratios produced by these logistic regressions. If there is less acquiescence in the incentive

(38% and 28%, respectively; $\chi 2(1)=4.35$; $p=0.04$) – with no difference by experimental condition for the income and employment records request (27% and 26%, respectively; $\chi 2(1)=0.11$ $p=0.74$).

group, the odds ratios should be significantly less than one. While five of the eight odds ratios were less than one, none of the results were significant. Overall, the incentive seems to have had minimal effect on acquiescence.

Table 2.9. Odds Ratios from Logistic Regressions Predicting Acquiescence Control=Ref)

	Item 15 *Odds Ratio (95% CI)*	**Item 16** *Odds Ratio (95% CI)*	**Item 23** *Odds Ratio (95% CI)*	**Item 24** *Odds Ratio (95% CI)*
Version A	0.93 (0.61, 1.44)	0.87 (0.60, 1.26)	0.74 (0.51, 1.08)	1.19 (0.81, 1.74)
Version B	0.86 (0.60, 1.23)	1.42 (0.95, 2.13)	1.01 (0.69, 1.47)	0.88 (0.61, 1.28)

Response order effects. In aural modes of data collection, such as telephone interviewing, both primacy and recency effects may occur (Krosnick & Alwin, 1987). To determine whether response order effects occurred, I calculated the proportion of the 27 items with at least four response options for which the respondent selected the first option, one of the first two options, the last option, or one of the last two options (Table 2.10). There was no evidence that the incentive had a significant impact on the proportion of items for which respondents selected one of the final options that were offered or one of the first options. I hypothesized that the incentive might have a greater effect on response order effects in the second half of the interview than in the first half; however, incentive group respondents were not less likely than control group ones to display response order effects in either half of the interview. Only one of these items was behavioral, so I did not look separately at the effect on attitudinal and behavioral items.

Table 2.10. Proportion of Items for Which Respondents Displayed Response Order Effects, by Incentive Condition

	Incentive *(Percent)*	**Control** *(Percent)*	***t*-value**
Primacy			
One of first two options	58.8	58.7	0.08
First option	29.2	30.4	-1.51
Recency			
One of last two options	29.9	29.5	0.71
Last option	16.9	17.2	-0.45
Sample size	(524)	(376)	

I also looked at the responses to each of the 27 items individually to determine whether there were particular items for which incentive group respondents were more likely to display primacy or recency effects. Overall, this analysis revealed very few significant differences between the incentive and control group responses. There were significant differences in the proportion of respondents selecting either the last response or one of the last two options for three items; in two of these three instances the control group respondents were significantly more likely to select the later options. There also were significant differences in the proportion of respondents selecting either the first option or one of the first two options for three items; again, in two of these three instances the control group respondents were significantly more likely to select the earlier options.

The weakness of these approaches is that they make it difficult to isolate response order effects from true differences in the attitudes and behaviors of respondents in the two groups. To control for this, I included response order experiments for four survey items. For each of these items, respondents were randomly assigned to one of two response option orders – half of the respondents heard the response options in the original

order and the other half heard them in the reverse order. For each of the four items, I calculated the percentage of respondents in the incentive and control groups that selected the two options that were presented first in the original order and last in the reverse order.

Table 2.11. Proportion of Respondents Selecting First Two Options in Original Order, by Incentive Condition

	Original Order: First Options *(Percent)*	**Reverse Order: Last Options** *(Percent)*	**Difference Between First / Last**
Health rating			
Incentive	47.5	32.3	15.2**
Control	43.4	35.6	7.6
Health insurance rating			
Incentive	57.7	47.9	9.8*
Control	58.8	42.6	12.2**
Most important for health			
Incentive	50.0	41.6	8.4
Control	49.4	47.9	1.5
Biggest problem for Congress			
Incentive	48.0	57.6	-9.6*
Control	48.8	55.8	-7.0

* $p<0.05$; ** $p<0.01$

As shown in the final column of Table 2.11, for the two rating questions, respondents were more likely to pick these two options when they were presented first than when they were presented last. The other two items presented respondents with a list of options and asked them to pick the one they felt was most important. For the first item there were not significant response order effects in the incentive or the control group, and for the second item, there was evidence of recency effects for the incentive group. For three of the four items, the difference between the responses obtained in the two presentation orders was larger in the incentive group than it was in the control group.

To determine whether the response order differences were *significantly* larger in the incentive group than in the control group, I estimated a logistic regression model for each item. The dependent variable was whether or not the respondent selected the two

options that were presented first in the original response order. The independent variables were whether or not the respondent was in the incentive group, whether or not he or she heard the response options in the original order, and the interaction of these two variables. As shown in Table 2.12, the interaction term was not significant for any of the four items. I also repeated this analysis looking at the proportion of respondents who selected the *last* two options in the original order and came to the same conclusion that there was not a significant interaction between incentive and response order.

Table 2.12. Logistic Regressions Predicting Selection of First Two Options in Original Order

	Health rating	**Health insurance rating**	**Most important for health**	**Biggest problem for Congress**
Intercept	-0.43**	0.48	-0.17*	-0.02
Incentive	0.01	0.05	-0.02	0.04
Original Order	0.24**	0.26**	0.09	-0.15*
Incentive*Original Order	0.08	-0.06	0.09	-0.02

* $p<0.05$; ** $p<0.01$

Use of round or prototypical numbers. The questionnaire included 15 items for which respondents were asked to provide open-ended numerical responses. Respondents may choose to estimate their response to such items instead of making the effort needed to come up with an exact answer. This decision would manifest itself in the use of round or prototypical numbers, such as five or ten. For items asking about the past week or a typical week multiples of seven could also be considered prototypical, while multiples of twelve also could be considered prototypical for items asking about the past year.

Virtually all of the respondents (over 99%) provided at least one such response. The proportion of respondents providing a round or prototypical response for a particular

item ranged from 19% (number of times saw a doctor in the past year) to 83% (number of hours worked each week). As shown in Table 2.13, there was not a significant difference in the mean proportion of items to which respondents provided a round or prototypical response between the incentive and control groups (47.4% and 46.7%, respectively). Respondents in both groups also were equally likely to provide each of the types of round or prototypical responses.

Table 2.13. Mean Proportion Items for which Respondents Provided Round or Prototypical Responses, by Incentive Condition

	Incentive *(Percent)*	**Control** *(Percent)*	***t*-value**
All varieties	47.4	46.7	0.46
Multiples of 5	36.5	37.1	-0.50
Multiples of 10	22.5	21.6	0.97
Other (7, 12)	16.2	14.6	1.45
Sample size	(524)	(376)	

Reviewing the responses to each item individually demonstrated that there was not a clear pattern to the results, with the incentive group being more likely to provide a round response for seven of the items and the control group being more likely to do so for the rest. There were significant differences between the incentive and control groups for two of the items. Respondents in the incentive group were significantly more likely than those in the control group to provide a round response for an open-ended question asking about their 2010 household income (54.3% and 44.8%, respectively; $\chi^2(1)=4.06$; $p=0.04$). Conversely, they were significantly less likely to provide a round response when asked how many years they had lived in their current home (27.7% and 34.0%, respectively; $\chi^2(1)=4.20$; $p=0.04$).

I hypothesized that the incentive might have a greater effect on respondents' decision to provide a round number in the second half of the questionnaire than it would

in the first half of the questionnaire. However, there was not a significant difference in the mean proportion of items to which respondents provided a round response in either half of the interview. It appears that the incentive had little impact on the likelihood that respondents provided round or prototypical numbers.

Estimation strategy for behavioral frequency items. After answering a question about how many times they had seen a doctor in 2010, respondents who reported at least one visit were asked to indicate how they came up with their answer: by thinking about each visit, by thinking about how often they usually go to the doctor, by thinking about types of visits, or by estimating based on a general impression. These final three options were considered to be estimation strategies. Respondents were permitted to select multiple responses.

Table 2.14 shows the proportion of respondents in each incentive condition that indicated having used each response strategy. Just over half of the respondents in the incentive and control groups respondents reported using at least one estimation strategy (55.4% and 51.7%, respectively). The incentive did not significantly impact the likelihood that respondents used at least one estimation strategy or that they used any particular strategy.

Table 2.14. Proportion of Respondents Reporting Each Strategy, by Incentive Condition

	Incentive *(Percent)*	**Control** *(Percent)*	$\chi^2(1)$
Each visit	35.1	38.7	1.13
How often usually go	22.2	18.9	1.32
Types of visits	22.0	18.1	1.91
General impression	14.4	15.3	0.12
Don't know / Refuse [A]	11.6	10.5	0.28
At least one estimation strategy	55.4	51.7	1.11
Sample size [B]	(473)	(354)	

[A] Repeating this analysis with "don't know" and "refuse" responses excluded did not change the outcome of the analysis.

[B] Respondents who did not answer the question about number of doctor visits or said that they had not seen a doctor were not asked this question.

Sixteen percent of respondents reported visiting a doctor only once in 2010. It seems likely that these respondents could easily recall this single visit and would be less likely to use an estimation strategy (half of them reported recalling each visit, as compared to about one-third of all respondents). Restricting the analysis to individuals who reported seeing a doctor at least twice indicated that incentive group respondents were marginally significantly more likely than the control group to use at least one estimation strategy (57.5% and 50.5%, respectively; $\chi^2(1)=3.20$; $p=0.07$).

Lack of attention to question wording. If respondents are taking cognitive shortcuts, they may not listen very carefully to the survey questions. To determine whether respondents were listening closely, four experiments were included in the questionnaire in which one half of the respondents in each incentive condition were told to exclude a category from consideration when answering the question. If respondents in the incentive group were paying greater attention than those in the control group, then the

difference between the two wording versions should be significantly greater among incentive respondents than among control respondents.

For all four items, the mean response provided in the exclusion condition was lower than that provided in the no-exclusion condition – this number was significantly lower for three of the four items.[10] For three of the four items the difference between the mean response in the exclusion and no-exclusion conditions was smaller in the control group than it was in the incentive group (Figure 2.3). An ANOVA was conducted for each item to determine whether there was a significant interaction between question wording condition and incentive condition. The interaction was not significant for any of the four items, suggesting that there was not a significant difference in attention to question wording between the incentive and control groups.

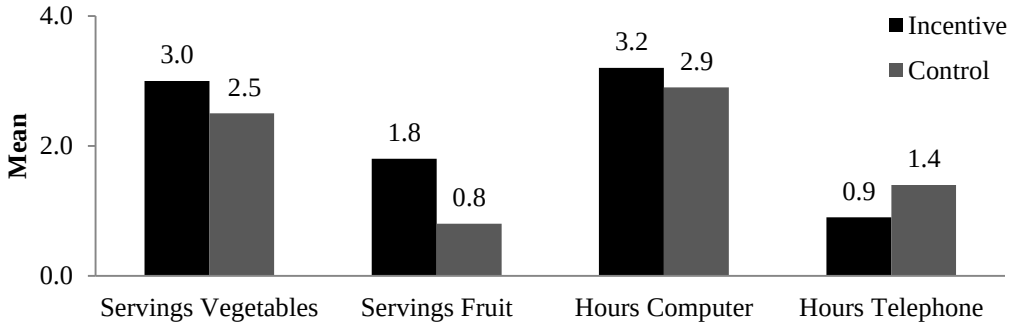

Figure 2.3. Difference between No-Exclusion Mean Response and Exclusion Mean Response, by Incentive Condition

Underreporting to filter items. For the filter items in this section, respondents were asked whether they had been diagnosed with a series of six medical conditions. Across the entire sample, the prevalence of each of the medical conditions varied to a

[10] The difference was not significant for the item asking about time spent on the telephone – this may have occurred because the exclusion category (time spent speaking to family members) was not a substantial enough proportion of the time respondents typically spend talking on the phone or because respondents did not report spending very much time speaking on the telephone even in the no-exclusion condition (mean=5.4 hours).

considerable degree, from 11% (arthritis) to 48% (hypertension). There were no significant differences in the prevalence of each of the conditions between the incentive and control groups. Respondents in both groups reported a similar mean number of conditions (1.4 in both; t=-0.27; p=0.79).

If respondents replied that they had been diagnosed with a particular condition, they were asked to indicate the age at diagnosis; if they replied negatively to the diagnosis item, they were asked whether they had been diagnosed with the next medical condition in the list. The order of the diagnosis questions was randomized. A logistic regression was estimated for each diagnosis item, with the dependent variable being whether or not the respondent provided an affirmative response to the item and the independent variables being position in the presentation order (1 through 6), whether or not the respondent received an incentive, and an interaction term. The interaction term was not significant for any of the six models (see Table 2.15), suggesting that receiving the incentive did not significantly affect the likelihood of underrreporting. Furthermore, there was not a clear pattern to the direction of the effect across the six items – for three items the interaction effect was positive and for three it was negative.

Table 2.15. Logistic Regressions Predicting Affirmative Response to Filter Items

	Diabetes	**Hyper-tension**	**Asthma**	**Arthritis**	**Heart disease**	**Anemia**
	(Estimate)	*(Estimate)*	*(Estimate)*	*(Estimate)*	*(Estimate)*	*(Estimate)*
Intercept	-1.28***	0.26	-2.17***	-0.56*	-1.80***	-2.06***
Incentive	-0.26	-0.29	-0.09	-0.13	0.23	0.08
Position in order	-0.09	-0.09	0.04	-0.05	0.07	-0.01
Incentive*Position	0.05	0.07	-0.04	0.08	-0.07	-0.04
Sample size	(900)	(900)	(900)	(900)	(900)	(900)

* p<0.05; ** p<0.01; *** p<0.001

In fact, it appears that underreporting did not occur in either experimental condition; estimating the logistic regression models without the interaction term showed that, controlling for experimental condition, position in the presentation order was not a significant predictor of providing an affirmative response for any of the items. There are two potential explanations for this result. First, the follow-up item may not have been burdensome enough to prompt respondents to underreport for the filter items; future research should examine the impact of incentives on underreporting in situations that pose greater burden to respondents. Second, only five questions were asked prior to the medical condition items; perhaps, if these questions had occurred later in the interview, more satisficing would have occurred.

Interview length. The total interview length ranged from 11.4 minutes to 55.8 minutes, with a median length of 19.5 minutes and a mean length of 21.0 minutes. Due to variation in the number of items each respondent was asked, I calculated the mean number of seconds per item asked and used this as a measure of interview length. As shown in Table 2.16, the mean number of seconds per item was significantly shorter in the incentive condition than it was in the control condition (16.1 and 16.7, respectively).[11] After controlling for age, this difference was only marginally significant ($p<0.07$). I hypothesized that that the effect of the incentive would be greater in the second half of the interview than it was in the first half. In the first half of the interview there was not a significant difference in the mean number of seconds per question between the incentive and control groups (17.1 and 17.6, respectively), while the mean number of seconds per

[11] This is equivalent to savings of about one minute per interview. Assuming a cost of $45 per interviewer hour, this translates into savings of about 75 cents per interview. This is much smaller than the savings described earlier ($6.08 per interview), suggesting that the reduction in call attempts was the main driver of the cost savings.

question was significantly shorter among incentive group respondents in the second half (15.3 and 16.0, respectively). However, a repeated-measures analysis indicated that the effect was not *significantly* greater in the second half than it was in the first half. Overall, these results suggest that the incentive led to short interview completion times, particularly in the second half of the interview.

Table 2.16. Mean Seconds per Question, by Incentive Condition

	Incentive	**Control**	***t*-value**
Full interview	16.1	16.7	2.03*
First half	17.1	17.6	1.47
Second half	15.3	16.0	2.34*
Sample size	(524)	(376)	

* $p<0.05$

Accuracy of responses as compared to frame data. For one of the survey items, length of residence in the respondent's current home, the information provided by respondents also was included on the frame. This is an item that respondents could potentially find burdensome and for which they might estimate their response; in fact, about one-third of respondents provided a round response for this item. Comparing respondent reports to the information on the frame provides an estimate of the accuracy of these reports.

The correlation between the number years that respondents reported living in their home and what was listed on the frame was 0.81. As shown in Table 2.17, there was not a significant difference in the proportion of respondents whose reports matched the frame exactly (21.7% and 19.5%, respectively) or in the mean difference between the respondent report and the frame value (4.2 and 5.0 years, respectively), suggesting that the incentive had no impact on the accuracy of these reports.

Table 2.17. Accuracy of Length of Residence Reports, by Incentive Condition

	Incentive	Control	Significance
Exactly the same as the frame	21.7%	19.5%	$\chi^2 = 0.63$
Mean absolute difference (years) [A]	4.7	5.0	$t = -0.61$
Sample size	(522)	(376)	

[A] This analysis was repeated with the differences trimmed at the 95th percentile. This did not change the non-significance of the result (4.2 and 4.5, respectively; $t = -0.78$).

Interviewer observation of effort. At the end of the survey, the interviewer was asked to make an observation about how much effort the respondent made during the interview. Table 2.18 shows the distribution of the ratings provided by the interviewers. Generally, interviewers provided favorable reviews of respondents' effort. Across both experimental conditions, they reported that about three-fourths of respondents answered the survey questions to the best of their ability "very often". There was not a significant difference between interviewers' impression of the effort put forth by the incentive and control groups

Table 2.18. Interviewer Ratings of How Often Respondent Answered Questions to Best of Ability, by Incentive Condition

	Incentive *(Percent)*	Control *(Percent)*
Very often	76.9	73.4
Pretty often	13.5	19.4
Somewhat often	5.7	4.5
Not that often	1.9	1.6
Not at all	1.9	1.1
	100.0	100.0
Sample size	(524)	(376)

$\chi^2(4)=6.80$; $p=0.15$; Note: Due to rounding, proportions may not sum to 100%.

Satisficing index. Finally, I determined whether the incentive affected the overall prevalence of satisficing behavior. To do so, I created an index of satisficing which measured the proportion of the items for which the respondent exhibited any satisficing behavior that could be identified for individual items at the respondent level – straight-

lining, providing a round response, recency, acquiescence, and item nonresponse. If a respondent straight-lined for a particular battery of items, they were considered to have satisficed for all items included in the battery. As shown in Table 2.19, on average respondents did one of these behaviors for about one third of the items they were asked. Respondents in the incentive and control groups committed these behaviors for a similar proportion of the items on average (33.4% and 33.9%, respectively). Again I hypothesized that the incentive might have a greater impact on satisficing behaviors in the second half of the questionnaire. However, there was not a significant difference between the incentive and control groups in either half of the survey. These results suggest that the incentive did not have a significant effect on the overall prevalence of these satisficing behaviors.

Table 2.19. Mean Proportion of Items for Which Respondent Satisficed, by Incentive Condition

	Incentive *(Percent)*	**Control** *(Percent)*	***t*-value**
Full interview	34.4	35.0	-1.39
First half	28.8	29.3	-0.95
Second half	39.1	39.6	-1.06
Sample size	(524)	(376)	

I also conducted a MANOVA with incentive condition as the independent variable and nine of the satisficing behaviors as the dependent variables.[12] As in the prior analysis, the multivariate test of differences suggests there was not a significant

[12] I included the following indicators: (1) length of response across all three open-ended items, (2) standard deviation of responses to batteries (non-differentiation), (3) number of affirmative responses (acquiescence), (4) proportion of items where respondent selected one of last two items (recency), (5) proportion of round responses, (6) deviation of length of residence report from frame value, (7) interviewer reports of respondent effort, (8) proportion of items skipped, and (9) average number of seconds per survey question. I excluded 'use of estimation strategy' because it was a dichotomous variable. I also excluded 'underreporting to filter items' and 'lack of attention to question wording' because these behaviors could only be identified in relation to other respondents – and not at the individual respondent level.

difference between the incentive and control groups in the prevalence of these satisficing behaviors ($F(4,215)=1.43$; $p=0.22$; Wilks' $\lambda=0.97$).

2.3.5 Variation in Incentive's Effect on Effort due to Respondent Characteristics

One limitation of the existing literature is that it often assumes incentives will have a similar effect on all of the respondents who receive them. However, it is reasonable to expect that the effect of the incentive on respondent effort will vary. In this section, I discuss the effect that cognitive ability, conscientiousness, and incentive recall had on the incentive's impact on respondent effort. I hypothesized that the incentive would have less of an effect on effort among respondents of lower cognitive ability. I also hypothesized that the greater the respondent's self-reported conscientious, the less the effect of the incentive would be. Finally, I hypothesized that the incentive would have little effect on the behavior of respondents who did not recall receiving it.

Distribution of respondent characteristics hypothesized to moderate incentive's effect on respondent effort: Cognitive ability. I examined three indicators of cognitive ability: age, education, and interviewer ratings. Respondents age 65 or older were considered to have lower cognitive ability than younger respondents. Respondents with a high school education or less were considered to have less cognitive ability than those with at least some college education. Respondents who were rated by the interviewer as having trouble answering the survey questions somewhat, pretty, or very often were considered to have lower cognitive ability. Table 2.20 shows the prevalence of these characteristics in the incentive and control groups. About half of the respondents were age 65 or older, one third had a high school education or less, and one in seven had

trouble understanding the survey questions. There were not significant differences between the incentive and control groups for any of the three indicators.

Table 2.20. Prevalence of Cognitive Ability Indicators, by Incentive Condition

	Incentive *(Percentage)*	**Control** *(Percentage)*	$\chi^2(1)$
Age 65 or older	48.3	51.1	0.65
High school education or less	33.1	30.8	0.53
Trouble answering questions	14.1	14.4	0.01

Conscientiousness. Respondents were asked ten items intended to measure the conscientiousness trait from the Big Five personality measure. I calculated the mean response for each respondent on a scale of one to five, with five representing the more conscientious end of the scale. I then determined the mean response in the incentive and control groups, and compared these values. The mean response in both groups was 4.2 (t=-0.39).

Incentive recall. At the end of the interview, respondents were asked two questions to determine whether they recalled receiving the incentive. As discussed earlier, respondents in the incentive group were significantly more likely than those in the control group to recall receiving the advance letter (86.5% and 66.0%, respectively). Among those incentive respondents who recalled receiving the letter, 94.7% recalled cash being included in the envelope. Two of the incentive respondents (less than 0.1%) recalled something else being in the envelope; one thought a brochure had been included and the other recalled "Dunkin' Donuts" being included.

As a point of reference, this follow-up question also was asked of control group respondents. Only one of them recalled cash being in the envelope (out of 376 respondents); this low incidence of false positives in the control group suggests that the incentive group respondents who said they remembered the incentive really did

remember it. Table 2.21 shows the distribution of recall of the advance letter and incentive among the incentive and control groups. Overall, most respondents' reports accurately reflected what they had received. Four-fifths of the incentive respondents recalled receiving the incentive (81.9%), while two-thirds of control group respondents recalled receiving an advance letter with nothing else in the envelope (66.0%).

Table 2.21. Advance Letter and Incentive Awareness, by Incentive Condition

	Incentive *(Percent)*	**Control** *(Percent)*
Recalled cash incentive	81.9	<0.1
Recalled something other than cash in envelope	0.4	0.0
Recalled letter but nothing else in envelope	4.2	66.0
Did not recall letter	13.5	34.0
Letter returned by Postal Service	0.0	<0.1
Sample size	(524)	(376)

Note: Due to rounding, proportions may not sum to 100%.

Variation in incentive effect across subgroups of respondents. I fit a linear model with satisficing index as the dependent variable and incentive, old age, low education, interviewer report of difficulty, and self-reported conscientiousness as the independent variables, as well as interactions between incentive receipt and each of these respondent characteristics.[13] The results of this analysis are shown in Table 2.22. As in the earlier analysis, there was no main effect of the incentive on the proportion of items for which the respondent satisficed; controlling for the other variables in the model, incentive group respondents satisficed for 35.4% of the items on average, while control group respondents did so for 35.6%.

[13] I also ran separate models for each characteristic in which the independent variables were the characteristic, incentive, and an interaction of these two variables. The results were largely the same, with the following exceptions: (1) conscientiousness had a marginally significant negative impact on satisficing (p=0.06) and (2) the interaction between age and incentive was not significant.

Two of the other variables did significantly predict satisficing. First, controlling for the other variables in the model, respondents age 65 or older satisficed for a significantly greater proportion of the items, on average, than did respondents under age 65 (37.2% and 33.8%, respectively). Second, controlling for the other variables in the model, respondents who had been rated by the interviewer as having more trouble answering the questions satisficed for a significantly greater proportion of the items, on average (36.9% and 34.1%, respectively). Both of these results are in line with what would be predicted by satisficing theory.

Table 2.22. Linear Models Predicting Respondent Effort

	Satisficing Index	**Proportion of Items Skipped [A]**	**Seconds per Question**
	Estimate (Standard Error)	*Estimate (Standard Error)*	*Estimate (Standard Error)*
Intercept	0.321** (0.020)	0.008 (0.014)	18.482** (1.637)
Incentive	0.031 (0.026)	0.006 (0.019)	-1.406 (2.132)
Age 65+	0.041** (0.006)	0.018** (0.004)	2.742** (0.457)
H.S. or less	0.004 (0.006)	0.003 (0.004)	-0.517 (0.488)
Iwr: trouble understanding	0.024** (0.008)	0.037** (0.006)	3.630** (0.654)
Conscientiousness average	0.000 (0.005)	0.001 (0.003)	-0.833* (0.364)
Age 65+ * incentive	-0.013+ (0.007)	-0.006 (0.005)	-0.196 (0.603)
H.S. or less * incentive	0.004 (0.008)	0.002 (0.006)	0.602 (0.635)
Iwr rating * incentive	0.007 (0.011)	-0.011 (0.008)	-0.239 (0.865)
Conscientiousness * incentive	-0.008 (0.006)	-0.002 (0.004)	0.173 (0.476)
Sample size	(867)	(867)	(867)
R-squared	0.172	0.145	0.213

+ $p<0.10$; * $p<0.05$; ** $p<0.01$

[A] This model was also estimated with a variable indicating whether or not the respondent skipped at least one item as the dependent variable. That model resulted in the same conclusions.

None of the interactions between the incentive and the characteristics of interest were significant at the 0.05 level. However, there was a marginally significant interaction between age and incentive ($p=0.08$); controlling for the other variables in the model, the incentive led to a slight decrease in satisficing among older respondents (from 37.7% to 36.8%) and had little effect on younger respondents (from 33.6% to 34.0%). Thus,

incentives may reduce somewhat the tendency of older respondents to satisfice more than younger ones.

I also conducted a MANOVA with nine of the satisficing behaviors as the dependent variables. The independent variables were the same as those used in the linear model predicting the satisficing index. This analysis produced very similar results to the linear model. The multivariate test of differences between the incentive and control groups was not significant ($F(9,843)=0.55$; $p=0.84$; Wilks' $\lambda=0.99$), and there were no significant interactions between the incentive and the other independent variables.

Next, I examined whether the effect of the incentive on the other data quality indicators differed according to respondent characteristics. Table 2.22 presents models looking at the two indicators for which I initially found significant incentive effects – item nonresponse and interview length. When respondent cognitive ability and conscientiousness were controlled for, the incentive no longer had a significant impact on these indicators, although the direction of the effect remained the same. There also were no significant interactions between the incentive and the characteristics of interest. However, several of the respondent characteristics did significantly affect respondent effort. As seen with the satisficing index, older respondents and those whom the interviewer rated as having more trouble understanding the questions were significantly more likely to display reduced effort. Also, respondents who were more conscientious answered significantly more quickly.

I estimated additional models predicting eight other effort indicators.[14] There was not a significant main effect for incentive in any of the models. Generally, the incentive

[14] I estimated one model for each indicator: linear models predicting (1) length of response across all three open-ended items, (2) standard deviation of responses to batteries (non-differentiation), (3) number of

did not interact with the respondent characteristics of interest, but there was one exception. In the model predicting the proportion of items for which respondents selected one of the last two response options, there was a significant interaction between incentive and the interviewer's judgment that the respondent was having difficulty answering the questions. The incentive resulted in a significant increase in recency effects among respondents of lower cognitive ability (from 26.6% to 31.2%; $p<0.01$) and had no effect among those of higher cognitive ability (from 30.1% to 29.7%; $p=0.60$). Overall, these results suggest that the effect of the incentive did not differ greatly according to respondents' cognitive ability or their level of conscientiousness.

Finally, I investigated the hypothesis that the incentive would have little effect on the behavior of respondents who did not recall receiving it. I began by comparing the mean satisficing index scores of incentive respondents who remembered the incentive, those who did not recall the incentive, and the control group respondents. Respondents who recalled the incentive satisficed for a significantly smaller proportion of the items than did control group respondents (34.1% and 35.0%, respectively; t = -2.31), while respondents who did not recall the incentive satisficed for a marginally significantly greater proportion of the items than control group respondents (36.1% and 35.0%, respectively; t = 1.67). The difference between the respondents who recalled the incentive and those who did not was significant at the 0.01 level as well.

Next, I estimated a linear model that also included the cognitive ability indicators (Table 2.23). In this model, respondents who recalled receiving the incentive still

---

affirmative responses (acquiescence), (4) proportion of items where respondent selected one of last two items (recency), (5) proportion of round responses, (6) deviation of length of residence report from frame value, and (7) interviewer reports of respondent effort. I also estimated a logistic regression model predicting (8) respondent use of an estimation strategy. It was not possible to estimate models for underreporting to filter items or lack of attention to question wording.

satisficed for a significantly smaller proportion of the items than did the control group. Respondents who did not recall the incentive satisficed for more of the items, but the difference between the two groups was no longer significant. As the magnitude of the effect was the same as that observed for the respondents who did recall the incentive, this lack of significance may be due to small sample size (429 incentive group respondents recalled the incentive, while only 95 did not).

Table 2.23 Linear Model Predicting Satisficing

	Satisficing Index
	Estimate *(Standard Error)*
Intercept	0.324** (0.003)
Recalled incentive (ref=control)	-0.008* (0.004)
Forgot incentive (ref=control)	0.009 (0.006)
Age 65+	0.036** (0.004)
H.S. or less	0.005 (0.004)
Iwr: trouble understanding	0.027** (0.005)
Sample size	(868)
R-squared	0.166

* $p<0.05$; ** $p<0.01$

Finally, I compared the three groups—those who recalled the incentive, those who did not recall the incentive, and the control group – for the same ten indicators included in the cognitive ability and conscientiousness analyses. Respondents who recalled the incentive differed significantly from control group respondents for only one of these ten effort indicators; they skipped a significantly smaller proportion of the items

on average than did the control group. Respondents who forgot the incentive differed significantly from control group respondents for four indicators – and in all four cases these differences suggested less effort among the respondents who were not aware of the incentive. Respondents who did not recall the incentive provided significantly fewer words on average in response to the open-ended items and displayed recency effects for a significantly greater proportion of the items on average. Also, the mean absolute discrepancy between self-reports and frame information for length of residence was significantly greater among incentive group respondents who were unaware of the incentive than it was among control group respondents. Finally, the interviewers rated these respondents' effort as being lower than the control group on average. I repeated these analyses with the three cognitive ability indicators as additional independent variables. Generally, the results were the same, with the exception that the difference in the interviewer ratings was no longer significant.

I also conducted a MANOVA with nine of the satisficing behaviors as dependent variables. This model also controlled for respondent cognitive ability. In this analysis, the multivariate test of differences between the control group and the incentive group respondents who were unaware of the incentive was significant ($F(9,846)=2.71$; $p<0.01$; Wilks' $\lambda=0.97$); the multivariate test of differences between the control group and incentive group respondents who recalled the incentive was not significant ($F(9,846)=1.56$; $p=0.14$; Wilks' $\lambda=0.98$). Overall, these results suggest that incentive group respondents who did not recall the incentive may have made less effort than the control group.

This raises the question of whether certain types of people were more or less likely to recall the incentive. For example, we might expect that respondents living in larger households were both less likely to see the advance letter mailing and more likely to face distractions while completing the survey interview. However, respondents living in larger houses were actually slightly *more* likely to recall the incentive, although this difference was not significant (84.3% and 80.8%, respectively; $\chi^2(1)=0.89$; $p=0.35$).[15] Additionally, we might predict that older respondents were both less likely to recall the incentive and more likely to satisfice. However, older respondents were actually slightly *more* likely to recall the incentive, although again this difference was not significant (84.1% and 79.8%, respectively; $\chi^2(1)=1.58$; $p=0.21$). A final possibility is that respondents who did not recall the incentive were less detail-oriented and generally less conscientious than those who did recall it; in fact, respondents who did not recall the incentive rated themselves as significantly less conscientious on average than did respondents who recalled the incentive (4.1 and 4.2, respectively; $p=0.04$)

[15] SSI included a variable on the frame indicating the likelihood that each sample member lived in a household of three or more people on a scale of 1 through 9. I considered respondents with a score of 7 or higher to be living in larger households.

2.4 DISCUSSION

Prepaid cash incentives are an effective tool for increasing survey response rates. However, their use can increase the cost per complete in telephone surveys. As a result, practitioners may need evidence of additional positive outcomes beyond increased response rates to justify their use. If incentives motivate certain respondents to participate who would not otherwise have done so, it is reasonable to believe that they may influence respondents' behavior *during* the survey interview. In particular, due to a sense of obligation to the researcher, incentives may increase respondent motivation during the interview, reducing the level of satisficing.

The existing evidence suggests that incentives may increase respondent effort in certain instances, but this effect is not always found. Additionally, the existing studies are limited in number and scope. The study described in this chapter aimed to overcome the limitations of the prior research investigating the effect of incentives on respondent effort. It examined the impact of a prepaid cash incentive on a wider array of effort indicators than has been used in the past. It also assessed whether the incentive's effect on effort depended on characteristics of the respondent or the item.

In the current study, the prepaid cash incentive led to a relatively large increase in the response rate (from 11% to 23%) and to a decrease in the cost per complete (by about $6). However, there was limited evidence that the incentive affected the level of effort that respondents made during the interview, suggesting that incentives may not affect respondent motivations beyond the point of their agreeing to take part in the interview. There were significant differences between incentive and control group respondents' behavior for only two of the twelve effort indicators included in the study.

First, the incentive led to a significant reduction in item nonresponse. This result provides some evidence of increased effort among incentive respondents. Given the lack of an effect on the other effort indicators, we can assume that these additional responses were of a quality comparable to those provided by the control group. In addition, respondents in the incentive condition completed the interview more quickly than those in the control condition. This result typically would be an indication of reduced effort among incentive respondents; however, given the lack of an effect on the other effort indicators, it may possibly be viewed as a positive outcome, as shorter interviews of equal quality mean reduced data collection costs, perhaps helping to offset the cost of the incentives.

There were slight improvements in effort among incentive group respondents for a few of the other indicators (e.g., slightly less acquiescence for most items), but these differences were small and not significant. Additionally, respondents in the control group were very unlikely to display some of the behaviors, such as underreporting to filter items or straight-lining; in such cases there was little opportunity for the incentive to result in improved effort.

I hypothesized that the incentive would have a greater impact in the second half of the interview than in the first half. For the two indicators where the incentive significantly affected effort – item nonresponse and interview length – a significant effect was observed in the second half of the interview and not in the first half; however, the effect was not significantly greater in the first half than in the second half for either indicator. I also hypothesized that the incentive would have a greater effect for attitudinal items than it did for behavioral ones; for the one effort indicator where I was able to test

this hypothesis, the incentive did not have a significant effect on responses to either type of item.

Finally, I hypothesized that the incentive would have a greater effect on certain respondents – those with greater cognitive ability, those who were less conscientious, and those who actually recalled receiving the incentive. The effect of the incentive did not differ according to cognitive ability or self-reported conscientiousness. There is some evidence that respondents who did not recall the incentive satisficed more often than the control group. Further research is necessary to understand which respondents are most likely to be affected by incentives and for what type of items we can expect to observe this effect.

Overall, these results should be reassuring for survey practitioners considering the use of prepaid cash incentives – there is no reason to believe that data quality will suffer as a result of this decision. However, these results do not suggest that prepaid cash incentives lead to much in the way of improved data quality. The limited improvements observed in the current study likely would be reduced even further in a typical RDD survey. There are two reasons for this. First, due to the difficulty of matching phone numbers to addresses, only about half of RDD sample members can be sent prepaid incentives – so any potential effects of the incentive would be limited to about half of the respondents. Second, the current study sent the advance letter to a specific individual and then interviewed that particular person; this design was ideal for the this study in order to maximize the effect of the incentive, but it may not be feasible or desirable for typical RDD surveys where sample member names may not be available or where researchers prefer to select a household member at random. A limitation of this study is that the

majority of the effort indicators it included were indirect measures of measurement error. To get a better understanding of the effect of incentives on measurement error, it would be useful to compare survey responses with validation data. I do this for several items in Chapter 3 and urge other researchers to do so whenever possible.

CHAPTER 3

THE EFFECT OF PREPAID CASH INCENTIVES ON RESPONSES TO SENSITIVE ITEMS

3.1 INTRODUCTION

Prepaid cash incentives are consistently found to increase mail survey response rates (e.g., Church, 1993; Edwards et al., 2002). However, researchers have expressed concerns that their use may have unintended effects on responses to survey questions. The current literature investigating the effect of incentives on response distributions finds limited support for this concern (Cantor et al., 2008; Singer & Ye, forthcoming). However, such studies are rather limited in number. Furthermore, when researchers have investigated the impact of incentives on survey responses, they typically have analyzed all of the survey items as one group. It is possible that the effect of the incentive varies depending on the type of item, and that the decision to analyze all of the items at once masks significant differences for subgroups of items. For example, responses to sensitive items appear to be subject to situational factors and survey design features (Tourangeau et al., 2000; Tourangeau & Yan, 2007); as a result, these items may be more susceptible than non-sensitive ones to incentive effects. Research repeatedly shows that respondents misreport for sensitive items (e.g., Magura & Kang, 1997; Tourangeau & Yan, 2007), so it would be useful to know whether incentives affect how honestly respondents answer such items.

This chapter describes the methods and results of an experiment using a prepaid cash incentive in a mail survey. It aims to extend the existing literature by determining the effect of incentives on a particular type of item – sensitive items. Furthermore, this study makes use of administrative data to determine the effect of the incentive on

accuracy at the individual respondent level and to quantify the resulting effects on nonresponse and measurement biases. Finally, this chapter considers the possibility that the effect of the incentive also varies by respondent characteristics.

3.1.1 Responses to Sensitive Questions

Research has shown repeatedly that respondents misreport when they respond to questions about sensitive behaviors and attitudes. For example, respondents have been found to underreport abortions (Tourangeau, Rasinski, Jobe, Smith, & Pratt, 1997), illicit drug use (Magura & Kang, 1997), and heavy drinking (Tourangeau & Yan, 2007), and to overreport voting (Belli, Traugott, & Beckmann, 2001) and religious service attendance (Presser & Stinson, 1998). Three main reasons for such misreporting have been proposed (Tourangeau & Yan, 2007). First, respondents may feel that a question is too personal or private – examples of this type of question include income and religion questions. Second, respondents may misreport because they are worried about the confidentiality of their responses; for example, respondents may lie about drug use for fear of legal consequences. Finally, respondents may misreport for questions that ask about behaviors or attitudes related to social norms, such as voting, because they want to avoid disapproval.

Some have argued that misreporting for sensitive questions is a personality trait (Crowne & Marlowe, 1964), while others suggest that it is susceptible to situational factors (Paulhus, 2002). In fact, survey researchers find that different question characteristics, such as the topic (Sudman & Bradburn, 1974; Bradburn, Sudman, & Associates, 1979) or reference period (Tourangeau et al., 2000), appear to affect the perceived sensitivity of items. They also find that certain design features, such as the use

of self-administered modes (as compared to interviewer-administered ones), may influence the accuracy of reports to sensitive questions (Kreuter, Presser, & Tourangeau, 2008; Tourangeau et al., 2000).

Receiving an incentive is another design feature that could affect misreporting for sensitive questions. In support of this hypothesis, in a review of laboratory experiments that varied the amount of monetary reward offered to participants, Camerer and Hogarth (1999) found that monetary incentives reduced presentation concerns; for example, self-reports of generosity and risk-preferring behavior were reduced when larger incentives were offered. The effect of incentives on such concerns may warrant further investigation in the context of surveys.

3.1.2 Theoretical Expectations for Effect of Incentives on Respondent Honesty

Several theories have been offered to explain why prepaid incentives increase response rates. As discussed in greater detail in Chapter 2, these theories suggest four possible effects of incentives on survey responses: (1) increased honesty due to respondents' sense of obligation to the survey researcher, (2) reduced honesty due to an influx of uninterested, unmotivated respondents, (3) reduced honesty due to a focus on extrinsic motivations instead of intrinsic ones, and (4) no effect beyond the point of agreeing to participate. Additional empirical evidence is needed to help determine which of these is the most accurate, and whether the effects of incentives vary according to survey design features or respondent characteristics.

3.1.3 Existing Studies Investigating Impact of Incentives on Sensitive Items

Typically, researchers have not observed a significant effect of incentives on response distributions (e.g., Shettle & Mooney, 1999; Willimack et al., 1995). However

there is some evidence that, at times, offering incentives can affect survey responses (e.g., Brehm, 1994; Singer et al., 2000). Usually, when researchers investigate the impact of incentives on survey responses, they analyze all of the survey items at once – and do not group them by characteristics such as perceived sensitivity. In their review of the incentives literature, Singer and Ye (forthcoming) comment that the lack of observed effects on response distributions may be due to researchers grouping all of the items together when conducting their analysis and not considering that the effect may differ across items.

Few existing studies have specifically investigated the impact of incentives on responses to sensitive questions. Medway et al. (2011) analyzed the effect of a prepaid cash incentive on responses to eight survey items that we might expect to be subject to social desirability biases. They found that the incentive led to a small increase in reporting of socially undesirable behaviors for almost all of the items (e.g., heavy alcohol consumption, receiving a speeding ticket, never reading the newspaper). However, the effect was significant for only one item: not having volunteered in the past year. This survey also included a mode experiment (sample members were randomly assigned to either mail or telephone administration); analyzing the data separately by mode suggested that the effect of the incentive on this item was limited to the mail survey. In a similar finding, utilizing a prepaid cash incentive resulted in a lower estimate of community involvement in a face-to-face survey (Groves et al., 2000).

A weakness of comparing response distributions is that, even when differences are observed between incentive and control group responses, it often is impossible to tell which group's responses exhibit less error. A second approach, which overcomes this

limitation, is to compare survey responses to validation data. In incentives research, this means that the relative accuracy of responses provided by those who have received an incentive can be compared to the accuracy of respondents who have not received one. However, using a record check to determine the accuracy of survey responses has rarely been done in conjunction with incentive experiments; in their review of the incentives literature, Singer and Ye (forthcoming) specifically point to the lack of research investigating the impact of incentives on the validity of survey responses.

I am aware of only two studies that have investigated the impact of incentives on responses to sensitive questions by comparing self-reports to administrative records. Krenzke et al. (2005) found that offering a promised cash incentive did not increase the accuracy of reports of recent drug use in the Alcohol and Drug Services Study; however, respondents were informed before the survey that they would be subject to a drug test. Tourangeau et al. (2010) also found that the use of a prepaid cash incentive did not have a significant impact on the accuracy of voting reports in a mail survey of registered voters, although it did lead to a small increase in misreports for both voting questions.

Finally, other than the responses themselves, there are indirect indicators of social desirability concerns. Respondents may refuse to respond to questions that they feel are too sensitive. For example, income questions are often subject to high rates of item nonresponse (Juster & Smith, 1997; Moore, Stinson, & Welniak, 1999). Medway et al. (2011) found that the use of a prepaid incentive reduced nonresponse to the income question in the mail version of the survey but had no effect in the phone version. Conversely, Tzamourani and Lynn (1999) found that providing a prepaid monetary incentive increased item nonresponse for sensitive questions. Krenzke and colleagues

(2005) found that offering a promised cash incentive did not have a significant impact on the proportion of items respondents skipped, regardless of item sensitivity.

As this review shows, the number of studies looking specifically at the effect of incentives on sensitive items is limited. In this chapter, I aim to increase our understanding of the effect of incentives on survey responses in three ways. First, because responses to sensitive questions appear to be sensitive to situational factors and survey design features, I consider the possibility that these items may be more susceptible to incentive effects than non-sensitive items. Second, for three survey items, I compare survey reports with record data to quantify the magnitude of reporting error. Finally, throughout the chapter, I consider the possibility that the effect of the incentive varies by respondent characteristics.

3.2 RESEARCH METHODS

3.2.1 Sampling Frame and Experimental Conditions

The data for this study come from a mail survey conducted in the fall of 2011 by Fors Marsh Group. The target population was Maryland residents who were registered to vote. Aristotle provided a sample of registered voters. This sampling frame was chosen because it included a name, address, and voting history for each case. Having full name information for each case allowed the survey materials to be targeted at a specific individual. Using this sampling frame also allowed me to compare self-reports of voting behavior with administrative records included in the Aristotle database; the file indicated whether each sample member had voted in the 2010, 2008, and 2004 general elections.

I utilized voting history information in my sample design because I felt that voting history would play a role in the effect of the incentive on the accuracy of respondent voting reports. Respondents are more prone to overreport voting than to underreport it, and overreporting is only possible among nonvoters; therefore, to determine the effect of the incentive on accuracy, it was important to have enough nonvoters in the sample for each election. Furthermore, I hypothesized that the effect of the incentive would be smaller for individuals who had voted occasionally in these three elections than it would be for individuals who had not voted in any of them; occasional voters are more likely than nonvoters to see voting as central to their self-concept, so receiving an incentive may not be sufficient to convince them to provide honest voting reports.

Thus, I included three groups in my sample: (1) “consistent voters” who had voted in all three elections, (2) “occasional voters” who had voted in one or two of the

elections, and (3) "consistent nonvoters" who had not voted in any of them. I aimed to yield a sample where one-third of the respondents fell into each of these groups. Using information from Tourangeau et al. (2010) concerning the impact of voting history and incentive receipt on response rates to a similar mail questionnaire, I developed the sample design shown in Table 3.1.

Table 3.1. Sample Design

	Total	**Incentive**	**Control**
Consistent voters	858	242	616
Occasional voters	1,170	330	840
Consistent nonvoters	1,872	528	1,344
Total	3,900	1,100	2,800

All 3,900 sample members were sent an advance letter, initial questionnaire, and reminder/thank you postcard. Eleven-hundred of the sample members were randomly assigned to receive a $5 prepaid cash incentive with the initial questionnaire. Sample members who had not responded after the first three mailings also were sent a replacement questionnaire. The field closed approximately three weeks after the final questionnaire was mailed. Table 3.2 shows the mailing schedule. The survey materials are included in Appendix C.

Table 3.2. Mailing Schedule

Mail Piece	**Mail Date**
Advance letter	November 4, 2011
Initial questionnaire	November 7, 2011
Reminder/Thank-you postcard	November 14, 2011
Replacement questionnaire	November 28, 2011
Survey fielding close date	December 20, 2011

3.2.2 Questionnaire: Responses to Sensitive Questions

The questionnaire was designed to assess the effect of the prepaid incentive on respondents' answers to sensitive questions. It included several measures of respondent

willingness to respond honestly: (1) responses to the "impression management" factor of the Balanced Indicator of Desirable Reporting (BIDR), (2) reports of socially undesirable behaviors or attitudes, (3) item nonresponse, and (4) accuracy of voting reports as compared to administrative records. The questionnaire covered several topics, including health, leisure activities, and involvement in politics. A copy of the questionnaire is included in Appendix C.

Impression management. In order to measure the deliberate effort to present oneself in a flattering light, I included 10 items from the "impression management" factor of the Balanced Indicator of Desirable Reporting (BIDR) (Paulhus, 1984). These items measure the degree to which the respondent is making a conscious effort to make a favorable impression on others.[16] Examples include, "I have never dropped litter on the street," and, "I sometimes tell lies if I have to". The items utilized a four-point scale, from "not true at all" to "very true". I calculated the mean response for each respondent on a scale of one to four, with "four" representing greater concern for impression management and "one" representing less concern.

Undesirable behaviors and attitudes. The questionnaire included items of varying levels of sensitivity. For most of these items, I had to rely on the assumption that socially undesirable behaviors and attitudes are underreported and that increased reporting represents increased honesty. I recoded each variable to be dichotomous, where a value of "one" indicated admission of a socially undesirable behavior or attitude. I determined an overall score for each respondent by calculating the proportion of items for which he or she admitted to doing the socially undesirable behavior or holding the

[16] Following the argument by Tourangeau and Yan (2007) that respondents make a deliberate decision to misreport for sensitive items, I did not include items from the other factor, "self-deceptive enhancement". Those items measure whether respondents believe that they are responding honestly (but have inaccurate views of themselves).

socially undesirable attitude. This score was calculated as a proportion because of skip patterns that resulted in variation in the number of items respondents were asked. Additionally, I only considered respondents to be eligible for each of the three voting items if they were at least 18 years old in the year of the election. I did not include responses to demographic items or the impression management items in this analysis.

Item nonresponse. There may be a positive relationship between item sensitivity and item nonresponse (Tourangeau & Yan, 2007). Therefore I used item nonresponse as an additional indicator of respondent sensitivity concerns. I calculated the proportion of the items that each respondent skipped. This analysis included only the items that were part of the index of undesirable behaviors and attitudes described above.

Voting reports. The sampling frame indicated whether or not respondents voted in the 2010, 2008, and 2004 general elections. Although there is likely some error in the frame data, we can assume that these errors are equally present for the incentive and control groups due to the random assignment of sample members to experimental conditions. The questionnaire also included items asking respondents whether or not they voted in these elections. For these items, I compared self-reports to the frame data to assess the accuracy of the respondent reports. I only included respondents whose self-reported birth year matched the birth year provided on the Aristotle frame;[17] this was done to increase the likelihood that the intended respondent had filled out the survey, increasing the chance that discrepancies between voting self-reports and frame values were due to reporting error and not due to another household member filling out the survey. Respondents who did not answer the question were excluded from the analysis.

[17] Ninety-two percent of the respondents provided a birth year that matched the one provided by Aristotle.

Finally, I only included respondents in the analysis who were at least 18 years old in the year in which the election occurred.

3.2.3 Questionnaire: Indicators of Item Sensitivity

Particular topics are seen as more sensitive than others; for example, questions about drinking and gambling are considered less sensitive than those about the use of illegal drugs or sexual activity (Bradburn et al., 1979). I hypothesized that the effect of the incentive would be more apparent for sensitive items than for non-sensitive ones. I measured item sensitivity in several ways (1) respondent ratings, (2) item nonresponse, and (3) coder ratings.

Respondent ratings. At the end of the survey, respondents were asked, "To what extent do you think that people you know might falsely report their answers to questions on the following topics?" The topics included medical conditions, diet, exercise, mental health, alcohol use, smoking, drug use, volunteering, voting, income, and religious service attendance. Respondents used a three-point scale from "not at all" to "a lot". They were told that their responses to these questions would be used to improve future surveys. I calculated the average rating for each topic.

Item nonresponse. Because there may be a positive relationship between item sensitivity and item nonresponse (Tourangeau & Yan, 2007), I also determined the proportion of respondents that skipped each survey item.

Coder ratings. As a final indicator of item sensitivity, I asked 20 coders to rate the sensitivity of each item on a five-point scale from "not at all sensitive" to "very sensitive". I determined the average rating for each item.

3.3 RESULTS

3.3.1 Response Rate

As shown in Table 3.3, the overall response rate was 26.3% (AAPOR RR1). The response rate was significantly higher in the incentive condition (41.1%) than in the control condition (20.4%) ($\chi^2(1)=172.44$; $p<0.001$).

Table 3.3. Disposition Codes and Response Rate, by Incentive Condition

	Total	**Incentive**	**Control**
Returned questionnaire	1016	448	568
Eligible, no returned questionnaire	91	33	58
Unknown eligibility, no returned questionnaire	2761	609	2152
Not eligible[1]	32	10	22
Total	3900	1100	2800
Response Rate (RR1)	26.3%	41.1%	20.4%

1. These sample members were determined not to be Maryland residents during the fielding period.

3.3.2 Cost per Complete

The overall cost per complete was $31.36 (Table 3.4). The incentive resulted in a reduction in the cost per complete (from $32.45 to $29.98). This decrease in cost per complete largely was driven by a reduction in printing and postage costs due to the higher response rate in the incentive group. The control group required 8.9 pieces of mail to be sent per complete, while the incentive group only required 4.1 pieces of mail per complete. Additionally, 87% of control group respondents required a second questionnaire package, while only 70% of incentive group respondents required one.

Table 3.4. Costs (Dollars), by Incentive Condition

	Total	Incentive	Control
Printing/Postage	15,570.00	4,214.00	11,356.00
Incentives[1]	5,245.00	5,245.00	0.00
Data Entry	11,988.00	4,361.00	7,627.00
Total	32,803.00	13,821.00	18,983.00
Cost per complete	31.36	29.98	32.45

1. Excludes cost of incentives for cases that were undeliverable.

3.3.3 Sample Composition

As shown in Table 3.5, there were not any significant differences between the incentive and control groups in terms of gender, race/ethnicity, or education. However, the age distribution differed significantly in the two groups. Generally, the members of the incentive group were younger than the controls. Fifty-six percent of incentive respondents were less than 55 years old, while this was only true of 44% of the control group respondents.

Table 3.5. Demographics, by Incentive Condition

	Incentive (%)	Control (%)	Chi-square
Gender (n=1009)			0.00
Male	44.9	44.9	(*p*=0.98)
Female	55.1	55.1	
Age (n=996)			15.31
18-24	11.4	8.4	(*p*=0.01)
25-34	11.4	10.0	
35-44	13.3	9.8	
45-54	19.9	15.9	
55-64	20.1	23.1	
65+	23.8	32.7	
Race/ethnicity (n=1000)			1.51
White, non-Hispanic	73.3	75.8	(*p*=0.68)
Black, non-Hispanic	17.4	14.7	
Hispanic	1.8	2.2	
Other race, non-Hispanic	7.5	7.3	
Education (n=995)			8.35
Less than high school	6.6	5.8	(*p*=0.14)
High school	23.6	19.5	
Some college, no degree	26.8	24.5	
Associate's degree	7.3	6.1	
Bachelor's degree	19.8	22.5	
Graduate degree	15.9	21.6	

3.3.4 Respondent Honesty

For the first three dependent variables – impression management, reports of socially undesirable behaviors and attitudes, and item nonresponse – I took a similar analytic approach. First, I fit a linear model in which the independent variables were incentive receipt (with "control" as the reference group), age (six categories, with "65+" as the reference group), and an interaction term. I began with this model because of the significant difference in the age distribution of the incentive and control groups. Next, I fit a second linear model in which the independent variables were incentive receipt (with "control" as the reference group), age (six categories, with "65+" as the reference group), gender (with "female" as the reference group), education (three categories, with "high

school or less" as the reference group), race (with "non-White" as the reference group), and interactions between incentive receipt and these demographic variables. I estimated this second model to ensure that any effect of the incentive observed in the initial model was not due to any differences in respondent characteristics in the incentive and control groups. I also repeated these analyses with the largest values for the dependent variables trimmed at the 99th percentile; I only report on these analyses if they changed the direction or significance of the results. I took a different approach for analyzing the effect of the incentive on the accuracy of voting reports, which I describe in greater detail later in the Results section.

Impression management. A factor analysis confirmed that all 10 impression management items loaded positively on a single factor. There was not a significant main effect of the incentive in either model (Table 3.6). The fitted mean for both experimental groups in both models was 2.8 (on a scale of 1 to 4, with 4 representing the highest rating of impression management). There were not any significant interactions between incentive receipt and the demographic variables. The incentive appears to have had little effect on impression management concerns.

Table 3.6. Means from Linear Models Predicting Impression Management Scores

	Model 1	**Model 2**
Raw means		
Incentive	2.8	2.8
Control	2.9	2.9
Fitted means		
Incentive	2.8	2.8
Control	2.8	2.8
Sample size	990	969
F-value	7.94**	5.17**
R^2	0.08	0.09

* $p<0.05$; ** $p<0.01$

Undesirable behaviors and attitudes. Across all 45 items, there was not a significant main effect of the incentive on the proportion of items for which the respondent admitted to an undesirable attitude or behavior in either model (Table 3.7). Respondents in both groups admitted to the undesirable behavior or attitude for just over one-fourth of the items on average. There were not any significant interactions between the incentive and the demographic variables in either model. Overall, these results suggest that, across all 45 items, the incentive had minimal impact on reports of undesirable behaviors and attitudes.

Table 3.7. Means from Linear Models Predicting Proportion of Undesirable Responses

	Model 1	**Model 2**
Raw means		
Incentive	27.9%	27.8%
Control	27.0%	26.8%
Fitted means		
Incentive	27.7%	27.8%
Control	26.2%	27.2%
Sample size	996	975
F-value	1.93*	11.10**
R^2	0.02	0.18

* $p<0.05$; ** $p<0.01$

Item sensitivity. Based on the coder ratings of sensitivity, I grouped the 45 items into equal-sized high, medium, and low sensitivity groups.[18] The mean sensitivity ratings for these subgroups of items were 3.2, 2.3, and 1.6, respectively (on a scale of 1 to 5, with 5 representing the highest sensitivity rating). Respondents admitted to a significantly smaller proportion of highly sensitive behaviors and attitudes than of medium or low sensitivity ones (t=13.26 and t=13.18, respectively); there was not a significant difference in the proportion of medium and low sensitivity items endorsed by respondents (t=0.04).

[18] I provide information about the results of the three item sensitivity analyses in Appendix D, as well as my rationale for selecting coder ratings as the main indicator of item sensitivity.

I re-estimated the two linear models separately for each group of items. The fitted means from Model 2 are shown in Figure 3.1. In both sets of models, the incentive resulted in a significant increase in reporting of undesirable behaviors and attitudes in the high sensitivity group, but did not have a significant effect in the medium and low groups. However, a repeated-measures analysis indicated that the effect of the incentive on the highly sensitive items was not *significantly* greater than the effect on medium and low sensitivity items.

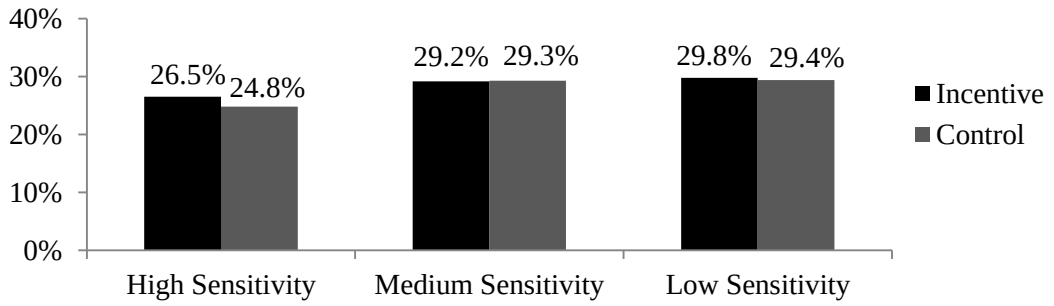

Figure 3.1. Mean Proportion Undesirable Behaviors and Attitudes, by Incentive Condition and Item Sensitivity

Item nonresponse. There was not a significant main effect of the incentive on the proportion of items skipped in either model (Table 3.8). Respondents in both groups skipped about four percent of the items on average. There also were no significant interactions between incentive receipt and any of the demographic variables. The incentive had little effect on the extent of item nonresponse.

Table 3.8. Means from Linear Models Predicting Proportion of Items Skipped

	Model 1	Model 2
Raw means		
Incentive	4.5%	4.5%
Control	4.6%	4.5%
Fitted means		
Incentive	3.6%	4.5%
Control	3.9%	4.4%
Sample size	996	975
F-value	9.75**	11.49**
R^2	0.10	0.19

* $p<0.05$; ** $p<0.01$

I conducted an analysis to determine whether the effect of the incentive on item nonresponse varied depending on the level of item sensitivity. Respondents did skip a significantly greater proportion of highly sensitive items than of medium or low sensitivity items (t=17.04 and t=11.61, respectively). However, the incentive did not appear to have a significant effect on item nonresponse for high, medium, or low sensitivity items (Figure 3.2).

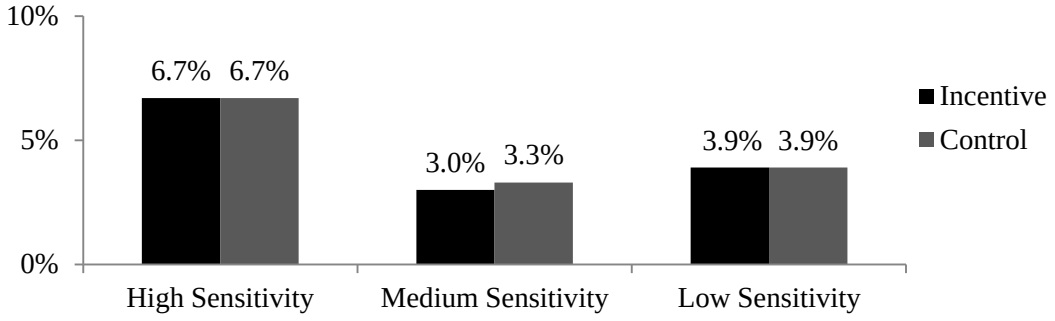

Figure 3.2. Mean Item Nonresponse, by Incentive Condition and Item Sensitivity

Voting reports. The survey included three questions (whether or not the respondent voted in the 2010, 2008, and 2004 general elections) for which information also was available on the frame. Thus, for these items I was able to determine whether or not individual respondents provided accurate survey reports. Figure 3.3 shows the

proportion of respondents in the incentive and control groups that provided inaccurate voting reports for each election. About one-fifth of respondents provided inaccurate responses for the 2010 and 2008 elections, while about one-fourth did so for the 2004 election. For all three items, incentive group respondents were more likely to provide inaccurate reports. These differences were not significant at the 0.05 level, though the difference was marginally significant for reports about 2008 (p=0.06).

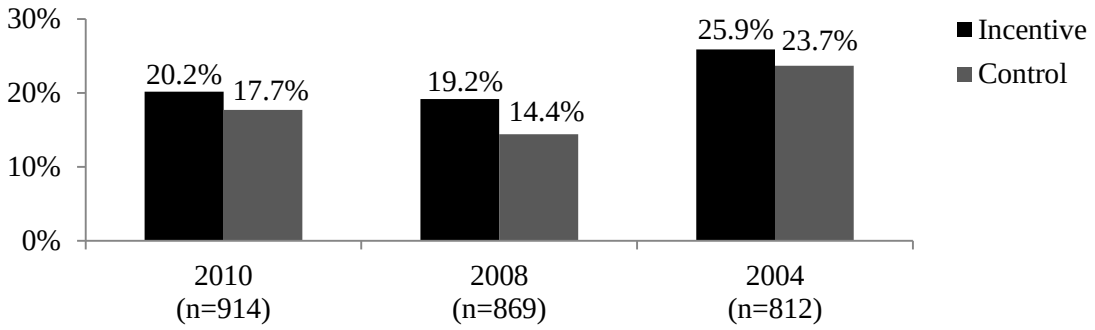

Figure 3.3. Proportion of Respondents that Provided Inaccurate Voting Reports, by Incentive Condition

Having both survey reports and frame data also allowed me to isolate the effect of the incentive on nonresponse bias from its effect on measurement bias. Table 3.9 provides information about the proportion of sample members that actually voted in each election; about one-fourth of the sample members voted in the 2010 election (28.0%), and about half of them voted in the 2008 and 2004 elections (43.0% and 53.0%, respectively). By design, there were not significant differences between the proportion of incentive and control group sample members that had voted in each election.

Table 3.9. Estimated Percentage of Voters and Bias Estimates, by Incentive Condition

	Entire Sample[1]	Respondents (Frame Value)	Respondents (Survey Response)	Nonresponse Bias	Measurement Bias
Voted in 2010					
Overall	28.0	45.3	57.3	17.3	12.0
Incentive	28.7	40.7	56.4	12.0	15.7
Control	27.7	48.9	58.1	21.2	9.2
Voted in 2008					
Overall	43.0	58.0	73.1	15.0	15.1
Incentive	43.8	54.3	70.7	10.5	16.4
Control	42.6	60.9	74.9	18.3	14.0
Voted in 2004					
Overall	50.0	59.5	73.0	9.5	13.5
Incentive	48.8	52.2	70.0	3.4	17.8
Control	50.5	65.0	75.3	14.5	10.3

1. To be comparable with the respondent values, this analysis excludes sample members who were under age 18 at the time of the election, respondents whose provided birth year did not match that listed on the frame, and respondents who did not answer the survey item.

I utilized this frame information to quantify the magnitude of the effect of the incentive on nonresponse and measurement bias for each of the three voting estimates. The magnitude of nonresponse bias was assessed by comparing the frame values for the full sample to frame values for the respondents, while the magnitude of measurement bias was assessed by comparing the respondents' frame values with their survey responses. As shown in Table 3.9, both the nonresponse and measurement biases led to overestimates of voting behavior. Generally, the magnitude of the two types of error was similar. For example, the survey results would have suggested that 73.1% of registered voters had voted in the 2008 election. However, only 43.0% of them had actually done so. This overestimates the prevalence of voting behavior by 30.1 percentage points – half of which was due to nonresponse bias (15.0%) and half of which was due to measurement bias (15.1%).

The statistical significance of the nonresponse bias for each item was assessed using the chi-square statistic resulting from comparisons of the proportion of respondents and nonrespondents that had voted in each election. The statistical significance of the measurement bias was assessed by conducting *t*-tests of dependent proportions that compared the proportion of respondents that reported voting in each election with the proportion that actually had voted. For the full sample there was significant nonresponse bias and significant measurement bias for all three elections; all of these biases were significant at the $p<0.0001$ level.

Incentive group respondents were less likely than control group respondents to have voted in each of the three elections; for example 65.0% of control group respondents voted in the 2004 election while only 52.2% of incentive group respondents

did so. These differences were significant at the 0.001 level for the 2010 and 2004 elections, and at the 0.10 level for the 2008 election.[19] The nonresponse biases generally remained significant at the $p<0.0001$ level when looking at the two experimental groups separately; the sole exception is that there was not significant nonresponse bias among incentive respondents for the 2004 election item. The nonresponse biases were smaller in the incentive group than in the control group for all three elections; for example, the nonresponse bias for the 2010 election was 21.2 percent in the control group and only 12.0 percent in the incentive group. However, the reduction in nonresponse bias was significant only for 2004 election ($p=0.003$).

Incentive group respondents also were less likely than control group respondents to report having voted in each election. For example, 60.9% of the control group respondents reported voting in the 2008 election, while only 54.3% of incentive group respondents did so. However, these differences were not significant at the 0.05 level (in 2004 the difference was marginally significant ($p=0.096$)). When looking at the two groups separately, the measurement biases remained significant at the $p<0.0001$ level for all three items. The measurement biases were larger in the incentive group than in the control group for all three elections; for example, the measurement bias for the 2010 election was 9.2 percent in the control group and 15.7 percent in the incentive group. However, the increase in measurement bias was not significant for any of the three items.

Voters versus nonvoters. Next, I considered the possibility that the effect of the incentive might differ depending on whether or not sample members had voted in each

[19] After controlling for age (due to differences in the age distribution of the incentive and control groups), the direction of the relationship remained the same, but the significance disappeared for the 2010 and 2008 elections. This suggests that the difference in the voting status between the incentive and control groups may have been driven by the influx of young respondents when the incentive was offered – younger people generally are less likely to vote (Belli et al., 2001).

election. For example, the effect of the incentive on nonresponse might be smaller among voters than among nonvoters because voters might be the type of people who are willing to participate in a survey even without having received an incentive. I began by looking at the effect of the incentive on the decision to participate separately for voters and nonvoters in each election. Figures 3.4 and 3.5 show that, regardless of incentive receipt, voters were more likely to participate in the survey than were nonvoters. Among both voters and nonvoters, the incentive led to a significant increase in the proportion of sample members who participated in the survey (all significant at $p<0.0001$).

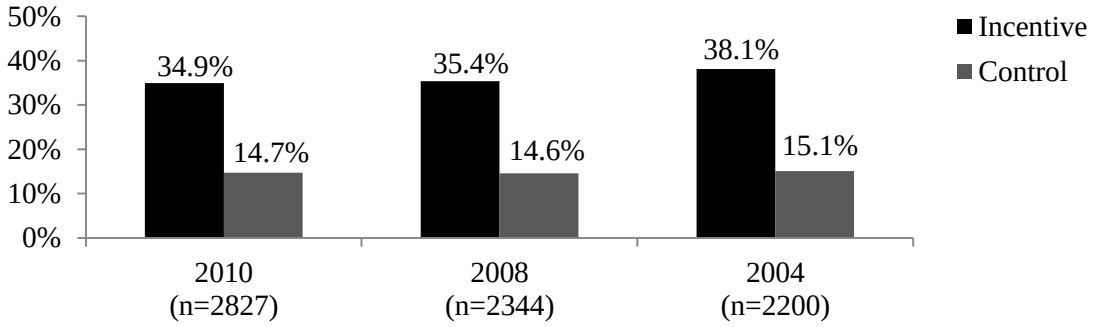

Figure 3.4. Proportion of Nonvoters Participating in Survey, by Incentive Condition. All differences between the incentive and control groups are significant at the 0.01 level or better.

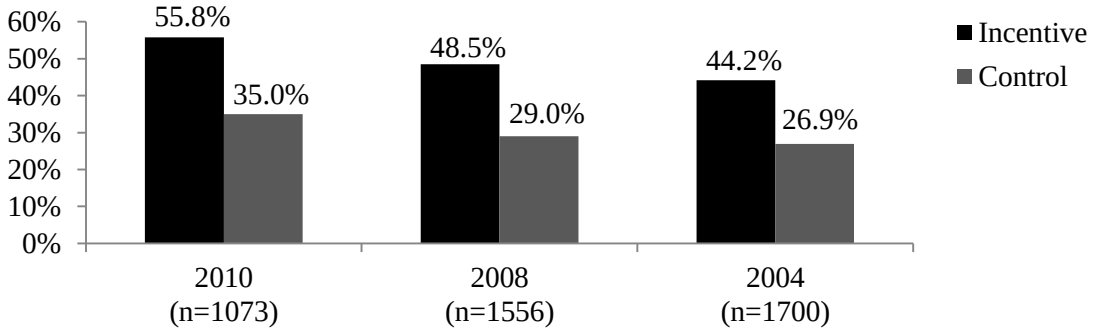

Figure 3.5. Proportion of Voters Participating in Survey, by Incentive Condition. All differences between the incentive and control groups are significant at the 0.01 level or better.

Next, I determined whether the incentive had a *significantly greater* effect on the participation decision among nonvoters than among voters. I estimated a logistic regression for each election in which the outcome was participating in the survey and the independent variables were incentive receipt, being a nonvoter, and an interaction term (Table 3.10). The interaction terms were positive and significant for two of the three models; the incentive had a significantly greater impact on the participation decision for nonvoters than it did for voters.

Table 3.10. Estimates from Logistic Regressions Predicting Survey Participation

	2010 election	**2008 election**	**2004 election**
Intercept	-0.62**	-0.90**	-1.00**
Incentive (ref=control)	0.85**	0.84**	0.77**
Nonvoter (ref=voter)	-1.13**	-0.87**	-0.72**
Incentive*Nonvoter	0.28	0.33*	0.47**
Odds Ratios: Incentive (ref=control)			
Voters	2.34	2.31	2.16
Nonvoters	3.10	3.22	3.46
Likelihood Ratio Chi-square	334.19**	266.19**	225.57**
Sample size	3900	3900	3900

* $p<0.05$; ** $p<0.01$

I also hypothesized that the effect of the incentive on respondent accuracy might be greater among nonvoters than among voters. Because voters have nothing to hide, they may provide accurate voting reports regardless of the incentive, while receiving the incentive might persuade nonvoters to report honestly. I again began by determining the proportion of respondents that provided inaccurate survey reports. As shown in Figure 3.6, inaccurate reports were relatively common among nonvoters; about one-third of nonvoters provided inaccurate reports for the 2010 and 2008 elections, and about one half did so for the 2004 election. The differences between the incentive and control groups in accuracy were very small and were not statistically significant.

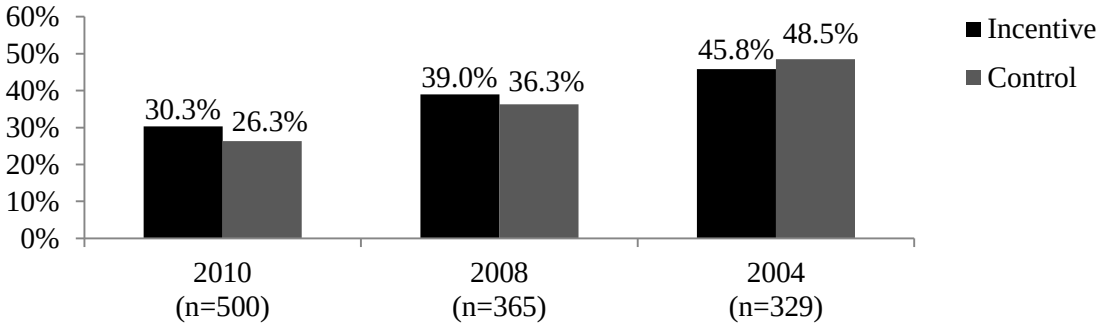

Figure 3.6. Proportion of Nonvoters Providing Inaccurate Voting Report, by Incentive Condition

As shown in Figure 3.7, inaccurate reports were much less common among voters, particularly for the 2008 election. Again differences in accuracy between the incentive and control groups were rather small and were not significant for the 2010 and 2004 elections. The incentive did lead to a significant increase in inaccuracy for the 2008 election (Fisher's exact test, $p<0.05$).

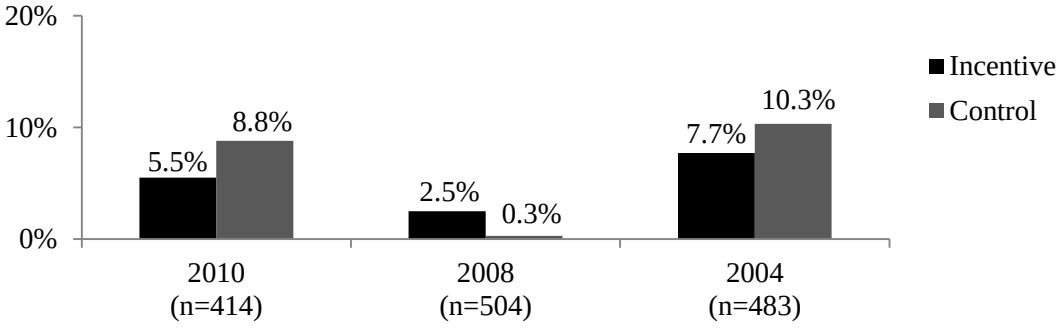

Figure 3.7. Proportion of Voters Providing Inaccurate Voting Report, by Incentive Condition

Next, I determined whether the incentive had a *significantly greater* effect on accuracy among nonvoters than among voters. I estimated a logistic regression for each election in which the outcome was providing an inaccurate voting report and the

independent variables were incentive receipt, being a nonvoter, and an interaction term (Table 3.11). None of the interaction terms were significant, suggesting that the effect of the incentive on accuracy did not vary significantly by voting status.

Table 3.11. Estimates from Logistic Regressions Predicting Inaccurate Voting Reports

	2010 election	**2008 election**	**2004 election**
Intercept	-2.34**	-5.70**	-2.17**
Incentive	-0.50	2.11	-0.31
Nonvoter	1.31**	5.14**	2.11**
Incentive*Nonvoter	0.69	-1.90	0.20
Odds Ratio: Incentive (ref=control)			
Voters	0.61	8.25	0.73
Nonvoters	1.21	1.23	0.90
Likelihood Ratio Chi-square	71.34**	233.99**	153.43**
Sample size	914	869	812

* $p<0.05$; ** $p<0.01$

Consistent voters, occasional voters, and consistent nonvoters. Finally, I considered the possibility that the effect of the incentive might differ depending on respondents' general pattern of voting behavior over the three elections. For this analysis I compared the effect of the incentive on respondents who consistently voted, occasionally voted, and consistently did not vote. I hypothesized that the less often a group of respondents had voted, the greater the effect of the incentive would be on their participation decision. As shown in Figure 3.8, the incentive significantly increased the proportion of sample members participating in the survey for all three groups ($p<0.001$ for all three groups).

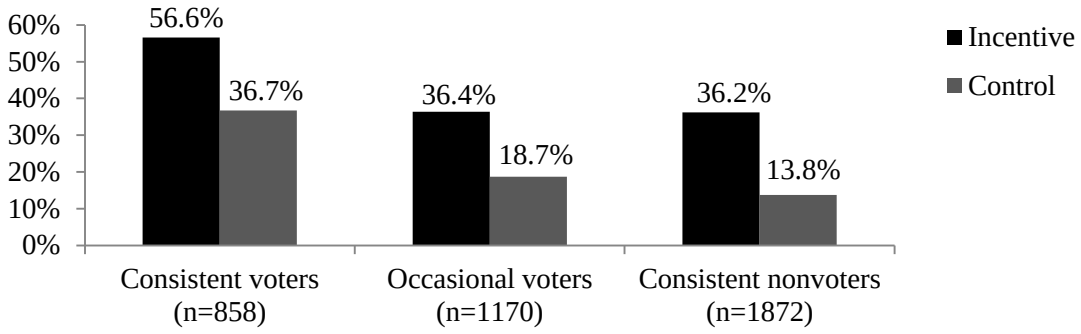

Figure 3.8. Proportion of Sample Members that Participated, by Voting History and Incentive Condition. The effect of the incentive was significant ($p<0.01$) for all three groups.

Next, I determined whether the incentive effect differed for the three groups. I estimated a logistic regression model predicting survey participation in which the independent variables were incentive receipt, voting history, and interaction terms (Table 3.12). The incentive had a significantly greater effect on the participation decision among consistent nonvoters than it did among consistent voters. However, there was not a significant difference in the effect of the incentive between consistent voters and occasional voters.

Table 3.12. Estimates from Logistic Regression Predicting Survey Participation

	Estimate
Intercept	-0.55**
Incentive	0.81**
Voting History	
Consistent Voter (reference)	--
Occasional Voter	-0.92**
Consistent Nonvoter	-1.29**
Incentive*Occasional Voter	0.10
Incentive*Consistent Nonvoter	0.46*
Odds Ratio: Incentive (ref=control)	
Consistent Voters	2.24
Occasional Voters	2.48
Consistent Nonvoters	3.56
Likelihood Ratio Chi-square	323.07**
Sample size	3900

* $p<0.05$; ** $p<0.01$

I also hypothesized that the effect of the incentive on the accuracy of voting reports would vary depending on respondents' voting history. I predicted that the effect of the incentive would be smaller for consistent voters than it would be for the other two groups because consistent voters are unlikely to provide false reports regardless of incentive receipt. I also predicted that the effect of the incentive would be smaller for occasional voters than it would be for consistent nonvoters; occasional voters' periodic attempts at voting suggest that this activity may be more important to them than it is for consistent nonvoters, and thus, receiving an incentive may not be sufficient to convince this group to provide more accurate responses. As shown in Figures 3.9 through 3.11, the incentive generally had little impact on respondent accuracy. The sole exception was for the 2010 election, where the incentive actually led to a significant increase in misreports among the consistent nonvoters (Figure 3.9).

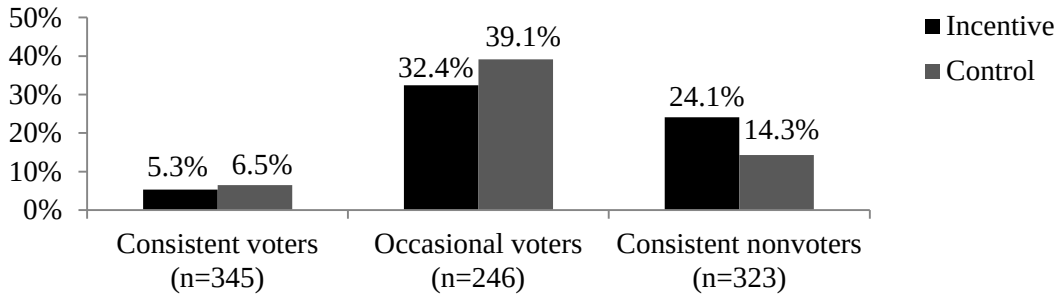

Figure 3.9. Proportion of Sample Members that Provided Inaccurate Voting Reports for 2010 Election, by Voting History and Incentive Condition

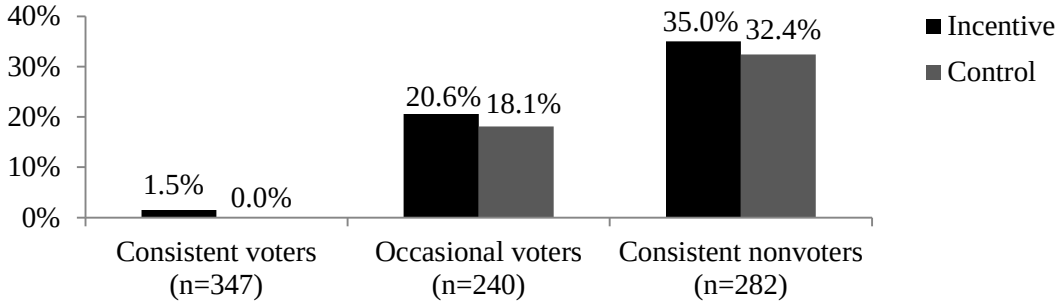

Figure 3.10. Proportion of Sample Members that Provided Inaccurate Voting Reports for 2008 Election, by Voting History and Incentive Condition

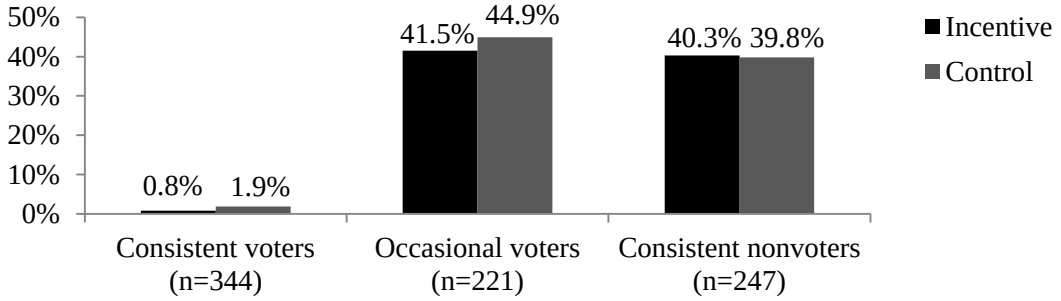

Figure 3.11. Proportion of Sample Members that Provided Inaccurate Voting Reports for 2004 Election, by Voting History and Incentive Condition

Next, I determined whether the effect of the incentive on respondent accuracy differed according to voting history. I estimated a logistic regression for each election item predicting inaccurate responses. In each model the independent variables were incentive receipt, voting history, and interaction terms (Table 3.13). For the 2010

election, the incentive had a significantly greater impact on inaccuracy among consistent nonvoters than it did among occasional voters. For the 2004 election, the effect of the incentive did not differ according to voting history. Due to poor model fit, the results from the 2008 model are not shown.

Table 3.13. Estimates from Logistic Regressions Predicting Inaccurate Voting Reports

	2010 election
Intercept	-2.66**
Incentive	-0.22
Voting History	
Consistent Voter (reference)	--
Occasional Voter	2.22**
Consistent Nonvoter	0.87*
Incentive*Occasional Voter	-0.07
Incentive*Consistent Nonvoter	0.86
Odds Ratio: Incentive (ref=control)	
Consistent Voters	0.80
Occasional Voters	0.75
Consistent Nonvoters	1.90
Likelihood Ratio Chi-square	94.26**
Sample size	914
	2004 election
Intercept	-3.97**
Incentive	-0.89
Voting History	
Consistent Voter	--
Occasional Voter	3.76**
Consistent Nonvoter	3.55**
Incentive*Occasional Voter	0.75
Incentive*Consistent Nonvoter	0.91
Odds Ratio: Incentive (ref=control)	
Consistent Voters	0.41
Occasional Voters	0.75
Consistent Nonvoters	1.02
Likelihood Ratio Chi-square	220.15**
Sample size	812

* $p<0.05$; ** $p<0.01$

MANOVA with all four outcomes of interest. Finally, I conducted a MANOVA with all four outcomes of interest (impression management score, proportion of socially

undesirable behaviors and attitudes, proportion of voting items with an inaccurate response, and proportion of items skipped) as the dependent variables. The independent variables were incentive receipt, age, gender, education, race, and interactions of incentive receipt with these demographic variables. As in the prior analyses, the multivariate test of differences suggests there was not a significant difference between the incentive and control groups in these outcomes ($F(4,879)=0.23$; $p=0.92$; Wilks' $\lambda=0.999$).

3.4 DISCUSSION

Extensive research has shown that prepaid cash incentives can lead to significant increases in mail survey response rates. In the current study, the prepaid cash incentive had several benefits. It led to a large increase in the response rate (from 20% to 41%), a small decrease in the cost per complete (by about $2.50), and improved representation of young people.

Researchers have raised concerns that using incentives may also have an effect on survey estimates, though existing studies find little support for this concern. The current study focused on a specific type of survey item – sensitive questions – to improve our understanding of incentives' effect on responses to such items in particular. Responses to sensitive items are affected by other features of the survey design (such as the mode of data collection), and these items also may be more susceptible than non-sensitive questions to incentive effects.

In the current study, there was little evidence that the incentive affected respondents' honesty in answering sensitive questions. There was not a significant effect on respondents' score on the Impression Management factor of the BIDR, the overall proportion of items for which respondents admitted to undesirable behaviors or attitudes, or the proportion of items that respondents declined to answer. Furthermore, the effect of the incentive did not differ according to respondent characteristics, such as age, education, race, and gender.

I hypothesized that the effect of the incentive would be greater for sensitive items than for non-sensitive ones. The incentive did lead to a significant increase in undesirable admissions among highly sensitive items (from 24.8% to 26.5%), while it did not have a

significant effect for items of moderate or low sensitivity. However, the effect of the incentive on the highly sensitive items was not significantly greater than the effect on medium and low sensitivity items.

For the three estimates of voting behavior in recent elections I was able to compare self-reports with frame data. This analysis demonstrated that there were considerable nonresponse and measurement biases, both of which led to overestimates of voting behavior. While the incentive resulted in a general pattern of reduced nonresponse bias and increased measurement bias for all three items, these effects generally were not significant. The incentive had minimal effect on the accuracy of responses. While the incentive appeared to have a significantly greater effect on the participation decision for nonvoters than for voters (particularly consistent voters), the effect of the incentive on accuracy did not vary significantly by voting history.

This study had a few limitations that should be addressed in future research. First, the sample members in this study were all registered voters. While this was necessary to be able to obtain voting records for the sample members, the response of this group to a prepaid incentive may not match that of the general population. Second, most of the measures of respondent honesty (with the exception of the accuracy of voting reports) were indirect indicators of measurement error; for example, we cannot be completely certain that the increase in highly sensitive admissions in the incentive group is not due to changes in sample composition. Furthermore, respondents considered "voting" to be the least sensitive of the 11 topics they were asked to rate; as the overall positive effect of the incentive was limited to highly sensitive items, it is possible that a significant impact on

accuracy would have been observed if I were able compare self-reports to validation data for more sensitive topics, such as drug use.

Overall, these results suggest that the incentive had little impact on respondent honesty. Still, this is encouraging news for researchers considering the use of prepaid incentives in mail surveys. There do not appear to have been any significant negative effects resulting from the use of the incentive.

CHAPTER 4

TESTING FOR MEASUREMENT INVARIANCE IN THE RESPONSES OF INCENTIVE AND CONTROL GROUP RESPONDENTS

4.1 INTRODUCTION

Prepaid cash incentives consistently lead to significant increases in survey response rates (Church, 1993; Edwards et al., 2002; Singer & Ye, forthcoming). However, researchers have expressed concern that paying respondents to participate in research might have unintended effects on the responses they provide to survey items. Typically, this possibility has been investigated by comparing the responses provided by respondents who have received an incentive with the responses of those who have not received one; for the most part, these studies have not found significant differences in the responses provided by the two groups (Cantor et al., 2008; Singer & Ye, forthcoming), and practitioners have used such findings to support the continued use of incentives. However, there are several drawbacks to using such an approach to address this research question; I describe these later. For groups of items that are intended to measure a latent characteristic, there is an available analytic method that may offer a better approach – testing for measurement invariance between responses provided by incentive and control group respondents. This method has not yet been applied to incentives research. This chapter extends the existing literature by conducting a measurement invariance analysis for four multi-item scales drawn from three recent surveys that included incentive experiments.

4.1.1 Limitations of the Existing Literature

There are several limitations to the typical approach of comparing incentive and control group responses. First, this method makes it impossible to differentiate between the incentive's potential effect on (1) sample composition and (2) how individuals respond to particular survey items. For example, incentive group and control group mean ratings of favorability toward the survey sponsor might appear to be quite similar. However, it could actually be the case that there are two concurrent forces at work that are cancelling each other out. First, the incentive might alter sample composition – it might attract respondents who have a weaker connection with the sponsor, thereby reducing favorability ratings. Second, the incentive might change the way that individuals respond to the items – it might cause respondents to feel obligated to provide more positive evaluations as a way of thanking the sponsor for the incentive, thereby increasing favorability ratings. Although the survey estimates of favorability would be identical in the two groups, the incentive group's responses would less accurately represent their true score on the underlying dimension of interest – satisfaction with the survey sponsor. If one group's responses are not a valid representation of the construct they are intended to measure, then it may not be appropriate to compare the two groups' responses (Osterlind & Everson, 2009). In their recent review of the incentives literature, Singer and Ye (forthcoming) specifically speak to the need for further research investigating the impact of incentives on the validity of survey responses.

Second, the typical approach of comparing response distributions assesses the effect of the incentive on responses to each item *in isolation* (e.g., "What proportion of respondents report being satisfied with the sponsor's customer support?"), without

considering that the relationship between responses to various items also may be altered (e.g., "Are respondents who are satisfied with customer support more likely to be satisfied with the quality of the sponsor's products as well?"). For example, similar proportions of incentive and control group respondents might respond positively to items on satisfaction with customer support and with products. However, if the incentive recruits respondents who are less interested in the survey and thus less motivated, they may put less thought into their respondents and may be generating their responses somewhat at random; this error may attenuate relationships between variables that are intended to measure the same construct – in this case, satisfaction with the survey sponsor. This issue also concerns the question of the effect of incentives on reliability – another area which Singer and Ye (forthcoming) argue is in need of further research.

4.1.2 Measurement Invariance

One method that overcomes these shortcomings is to test for measurement invariance. This approach is based on the assumption that particular survey items are intended to measure underlying, or latent, traits. For example, a survey might include several items whose purpose is to measure respondents' level of patriotism. Measurement invariance is said to exist if people from subgroups of interest – such as the incentive and control groups – who have the same amount of the latent trait (in this case "patriotism") also have the same probability of giving a particular response for a particular survey item (Lord, 1980; Mellenbergh, 1989). Such a result suggests that membership in this subgroup – in this case, the group that has received an incentive – is not a factor in determining the relationship between the observed response and the true value of the latent trait that the item is intended to measure (Millsap, 2011). Matching respondents on

the latent variable allows researchers to determine whether the incentive has led to differences in how well the survey items function across the two groups – and removes any issues about differences in the distribution of the latent trait.

If measurement invariance cannot be established for the responses of the incentive and control groups, this suggests that comparisons of the response distributions of the incentive and control groups are not appropriate (Vandenberg & Lance, 2000). There are two common approaches for assessing measurement invariance: (1) differential item functioning analysis (DIF) and (2) confirmatory factor analysis (CFA). Meade and Lautenschlager (2004) argue that researchers should conduct both approaches whenever possible.

4.1.3 Existing Studies Testing for Measurement Invariance

The application of measurement invariance methods in survey research has been somewhat limited to date. The most common application of such methods are efforts to demonstrate measurement invariance between web and paper versions of survey questionnaires (e.g., Cole, Bedeian, & Feild, 2006; Whitaker & McKinney, 2007) or between cultures in international surveys (e.g., Durvasula, Lysonski, & Watson, 2001; Kaminska et al., 2010; Welkenhuysen-Gybels, Billiet, & Cambré, 2003). For example, Davidov and colleagues used confirmatory factor analysis to establish measurement invariance across 20 countries in responses to the 2002-2003 European Social Survey (Davidov, Schmidt, & Schwartz, 2008).

In comparing the responses of respondents who have received an incentive with those who have not received one, researchers almost universally skip the step of demonstrating measurement invariance; I am not aware of any existing papers that have

tested for measurement invariance between incentive and control group respondents. If incentives influence respondents' approach to completing the survey interview – for example, by way of altered level of effort or self-presentation concerns – it is reasonable to expect that this may affect the relationship between the latent construct and the observed responses.

4.2. RESEARCH METHODS

4.2.1 Datasets

This chapter utilizes three survey datasets. Each of these surveys was selected because it met several criteria. First, each included an experiment in which some of the sample members were randomly assigned to receive a token cash incentive prior to the survey, while the other sample members did not receive an incentive. Second, each survey included at least one series of items that was intended to measure some latent trait; responses to these series of items were analyzed to test for measurement invariance.

JPSM Practicum survey. The first survey dataset came from the 2011 JPSM Practicum survey. As discussed in Chapter 2, this nationwide telephone survey utilized a listed sample. All sample members received an advance letter; 40% of the sample members were randomly assigned to receive a $5 bill with the advance letter, while the other sample members did not receive an incentive. The incentive resulted in a significant increase in the response rate, from 11% to 23%.

The survey included two series of items that were intended to measure a latent trait. The first was a series of ten items intended to measure patriotism. These items were based on questions asked in the 1996 and 2004 administrations of the General Social Survey. Respondents were asked how proud they are of various aspects of American society, such as scientific or economic achievements, on a four-point scale from "not proud at all" to "very proud". The second series of ten items was intended to measure conscientiousness, one of the personality traits that is part of the Big Five personality measure (Costa & McCrae, 1992). Respondents were asked the extent to which they agree that several statements accurately describe their personality on a five-point scale

from "strongly disagree" to "strongly agree". These items were adapted from a measure developed by Buchanan et al. (2005).

Maryland registered voter mail survey. The second dataset came from a 2011 mail survey of registered voters living in Maryland. As discussed in Chapter 3, all 3,900 sample members received an advance letter; 1,100 of them also received a $5 bill along with this letter. The incentive resulted in a significant increase in the response rate, from 26% to 41%. This survey included a series of ten items intended to measure impression management – the deliberate effort to present oneself in a flattering light; this is one of the two factors in the Balanced Indicator of Desirable Responding (Paulhus, 1984). Respondents were asked the degree to which certain statements described their behavior, such as, "I have never dropped litter on the street". The items utilized a four-point scale from "not true at all" to "very true".

Survey of Consumers. The final dataset came from the November-December 2003 and January-February 2004 administrations of the Survey of Consumers. This random-digit-dial telephone survey was conducted by the University of Michigan Survey Research Center. About half of the phone numbers could be matched to addresses. All of these cases were sent an advance letter and were randomly assigned to one of three experimental conditions: (1) $10 included with advance letter, (2) $5 included with advance letter, or (3) advance letter without an incentive. The interviewers were not told whether or not sample members had received a prepaid incentive.[20] Once contact was made, interviewers selected one of the adult household members at random to participate in the survey. Both the $10 and $5 incentives resulted in a significant increase in the

[20] Curtin et al. (2007) note that interviewers were permitted to offer refusal conversion incentives of either $25 or $50 at their discretion; typically this was done after two "hard" refusals. However, information about which specific respondents received these refusal conversion incentives was not available.

response rate as compared to the control group (67.7%, 63.8%, and 51.7%, respectively); however, there was not a significant difference between the response rates in the two incentive conditions (Curtin et al., 2007).

This survey included five items that were used to construct the Index of Consumer Sentiment. These items ask respondents their opinions about their, and the country's, current financial situation, as well as how this compares to years past and what they anticipate for the coming years. The exact wording of these questions is provided in Appendix E.

4.2.2 Identification of Eligible Cases

Measurement invariance analyses require complete data for each case. For each series of items, I identified respondents who had skipped at least one of the items and removed them from that analysis. For each survey, I also excluded respondents whose advance letter had been returned as undeliverable. There were two respondents of this type in the Practicum survey, and in the Survey of Consumers there were 72 such cases. I determined the proportion of respondents that remained eligible for analysis in each experimental condition for each series to ensure that these proportions were not significantly different from one another.

Limiting the analysis to cases for which there is complete data is rather common practice in the measurement invariance literature (e.g., Cole, 1999; Davidov et al., 2008; Levine et al., 2003; Scandura, Williams, & Hamilton, 2001). However, because this approach resulted in excluding almost one-fifth of the cases for the analysis of the patriotism items, I repeated all of the measurement invariance analyses with imputed values. I imputed for missing responses based on respondents' answers to the other items

in the series. I utilized IVEware to impute the responses separately for the incentive and control groups; this software estimates a series of multiple regression models in which the response to the item is predicted based on the other available variables – in this case, responses the other items in the series (Raghunathan, Solenberger, & Van Hoewyk, 2002). Imputing the missing values did not change the results of any of the analyses. I provide these results in Appendix F.

4.2.3 Comparison of Mean Responses

I began by calculating an index score for each series. For each set of items, I summed respondents' answers to each of the individual items.[21] Where necessary, I reverse coded variables so that all of the items in the series were scaled in the same direction (e.g., higher responses indicate more of the latent construct). For each series, I compared the mean index value for each experimental group and utilized *t*-statistics to identify statistically significant differences.

4.2.4 Internal Consistency Reliability

I began by calculating correlations between responses to each of the items in each series. I did this separately for the incentive and control groups. As a measure of internal consistency, I also calculated Cronbach's coefficient alpha for each series for experimental group (Cronbach, 1951). It is calculated as

$$\alpha = \frac{N \cdot \bar{c}}{\bar{v} + (N-1) \cdot \bar{c}},$$

where N is the number of items, $\bar{c}$ is the average inter-item covariance, and $\bar{v}$ is the average variance. This statistic indicates how closely the items are related, or how well

[21] This is a different approach than is taken for calculating the Index of Consumer Sentiment that typically is reported by the Survey of Consumers. That index is calculated to represent a single value for the full sample of each administration, as opposed to creating a single score for each respondent as I have done here. See http://www.sca.isr.umich.edu for additional information on how that index is calculated.

they measure a single construct; it can range from zero to one, with larger values indicating greater consistency. Typically, values above 0.70 are considered acceptable, and values above 0.90 are considered excellent (Kline, 2011). I determined whether the incentive and control group alpha values were significantly different from one another by calculating a *W*-statistic (Feldt, 1969; Feldt & Kim, 2006). This statistic is calculated as

$$W = \frac{(1-\hat{\alpha}_2)}{(1-\hat{\alpha}_1)},$$

where $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are the alpha values for the two groups. When the product of the sample size and the number of items in the series is greater than 1,000, the *W*-statistic follows an *F*-statistic distribution with degrees of freedom $v_1 = N_1 - 1$ and $v_2 = N_2 - 1$. If the *W*-statistic is greater than the critical *F*-value, then the two alpha values are significantly different from one another.

4.2.5 Differential Item Functioning

Differential item functioning (DIF) analysis tests whether, conditional on the latent score, the probability of providing a particular response to an item is the same in the subgroups of interest (Osterlind & Everson, 2009). DIF analysis emerges from an item response theory framework. This framework is based on two claims: (1) responses to an item can be predicted based on latent traits and (2) the relationship between the item response and the latent trait can be modeled with an item characteristic curve (Hambleton, Swaminathan, & Rogers, 1991). If this relationship differs in the subgroups that are being studied, then DIF is present.

To test for DIF, I used ordered logistic regression. Swaminathan and Rogers (1990) initially proposed the use of logistic regression models for identification of DIF in

binary items, and Zumbo (1999) extended this to propose the use of ordered logistic regression models for identification of DIF in polytomous items, such as the ones utilized in the current datasets. Although these methods have not been utilized as frequently as others, such as the Mantel-Haenzel procedure, simulation studies have shown that logistic regressions perform as well, or better than, these other methods in terms of identifying DIF (e.g., Narayanan & Swaminathan, 1996).

To test for DIF, I estimated two ordinal logistic regression models for each survey item. In each model, the dependent variable was the respondent's answer for that item. In the first model, the independent variable was the respondent's total score for that series of items; this model suggested that the probability of providing a particular response to that survey item depended solely on the respondent's total score on the latent variable. The second model also included a variable indicating whether or not the respondent received the incentive, as well as an interaction between total score and incentive receipt; this model suggested that the relationship between the item score and the latent score may differ according to whether or not the respondent received an incentive, and that this difference may vary depending on the respondent's latent score.

I recorded the likelihood-ratio chi-square value associated with the two models. Then, I calculated the difference between the two values. As the value from the first model had one degree of freedom, and the value from the second model had three degrees of freedom, the resulting difference value was compared to a chi-square distribution with two degrees of freedom (Millsap, 2011). A significant difference between the chi-square values indicated that the model fit was significantly improved by adding the additional parameters in the second model. For DIF to exist, Zumbo and Thomas (1997) suggest

that the resulting chi-square value should have a *p*-value less than or equal to 0.01; this more stringent value is suggested due to the large number of tests.

A benefit of utilizing the logistic regression approach to identifying DIF is the ability to estimate the magnitude of the effect (Crane et al., 2007). This was done by calculating the difference between the pseudo *R*-squared values produced by the two models (McKelvey & Zavoina, 1975; Zumbo, 1999). Values less than 0.035 indicated negligible DIF, while values between 0.035 and 0.070 indicated moderate DIF, and values greater than or equal to 0.070 indicated large DIF (Gelin & Zumbo, 2003). All DIF analyses were conducted in SAS.

4.2.6 Confirmatory Factor Analysis

As discussed earlier, I examined four series of items; the items in each set were indicators of a single latent variable. To test for measurement invariance, I estimated a series of one-factor measurement models for each set of items. At each step, constraints were added to the model to test for stronger forms of invariance. For each model, several model fit indices were reviewed; if the model fit the data acceptably well, then this form of invariance was considered to be established for this set of items.

In the first step, I began by estimating separate measurement models for the incentive and control groups. No constraints were placed on these models. I reviewed several model fit indices (discussed at greater length in the next section) to determine whether the measurement models were acceptable for each group of interest separately.

Next, I tested for configural invariance. In this step, I tested whether two things were consistent across the incentive and control groups: (1) the number of factors in the model and (2) which measured variables were indicators of each latent variable (Schmitt

& Kuljanin, 2008). In this step, a single measurement model was estimated for the two groups simultaneously. Two constraints were added to this model: (1) the number of factors and (2) the measured variables that were indicators of each latent variable. For the measurement models proposed in the current analysis, this simply meant that both groups were constrained to a single factor model. In this step, the actual factor loadings were allowed to differ between the two groups. Thus, separate estimates were produced for each group, but one set of model fit indices was produced for both groups simultaneously, and overall model fit was assessed again. If these models did not fit the data, this suggested that configural invariance did not exist.

Then, I tested for a stronger form of invariance – metric invariance. In this step, I tested whether the factor loadings were equivalent in the incentive and control groups (Schmitt & Kuljanin, 2008). Again, a single model was estimated for the two groups simultaneously, with the same constraints that the items were all indicators of a single factor. In this step, an additional constraint was added that the factor loadings for each item were equal for the incentive and control groups. To test whether this constraint held, the chi-square value produced by this model was compared to that produced by the model in the previous step. If the difference was not significant, this suggested that the addition of this constraint did not significantly worsen model fit, and that the equality constraint held.

If this constraint did not hold, I tested for partial measurement invariance – this suggests that invariance holds for some of the items in the model but not for others (Millsap, 2011). The purpose of this step was to identify the items responsible for the metric noninvariance identified in the previous step. To identify these items, it was

necessary to review the modification indices (MI). One MI was produced for each item; this value indicated the benefit in model fit (chi-square value) that would be achieved by removing the constraint that the loading for this item must be identical across the two groups. If none of the MI's indicated that removing particular constraints would lead to a significant change in model fit, then none of them were removed. However, if any of the MI's indicated that removing particular constraints would lead to a statistically significant improvement in model fit, then I released the constraint that had the largest impact on model fit. I then re-ran the model without this constraint and repeated the process of looking at overall model fit and the MI's until model fit was no longer significantly worse than what was observed for the configural invariance model. At this point, the parameters for which the constraints remained were considered to exhibit measurement invariance, and the parameters for which the constraints had been removed were considered to differ across the groups. All CFA analysis were conducted using the software package LISREL (Jöreskog & Sörbom, 2006).

Model fit indices. I assessed the global fit of each model by looking at several model fit indices: chi-square, the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMR), and the comparative fit index (CFI). Generally these statistics test whether there is a significant difference between the observed sample covariance matrix and the matrix produced by the estimated model parameters (Kline, 2011).

Chi-square is the traditional measure of model fit; it indicates the magnitude of the discrepancy between the observed sample covariance matrix and the covariance matrix produced by the estimated model parameters (Reise, Widaman, & Pugh, 1993). A

p-value greater than 0.05 suggests that the model is consistent with the observed covariance data. A weakness of the chi-square test is that, with large samples, it can be an overly sensitive fit test (Byrne, 1998; Kline, 2011).

The RMSEA overcomes this limitation by adjusting for sample size. It is calculated as

$$\sqrt{\frac{\chi_M^2 - df_M}{df_M (N-1)}},$$

where χ_M^2 is the model chi-square value, df_M is the degrees of freedom in the model, and N is the sample size (Kline, 2011; Steiger & Lind, 1980). Values below 0.07 indicate good fit (Steiger, 2007). Typically, the 90% confidence interval is provided along with the RMSEA value.

The SRMR is an indicator of the mean absolute correlation residual – or the overall difference between the observed and predicted correlations (Kline, 2011). The SRMR ranges from zero to one; values less than 0.05 indicate good fit (Byrne, 1998; Diamantopoulos & Siguaw, 2000), while values less than 0.08 indicate fair fit (Hu & Bentler, 1999).

Finally, the CFI determines the relative improvement in the fit of the proposed model over a baseline model – typically one where the measured variables are assumed to be uncorrelated or where there is no common factor (Millsap, 2011). It is calculated as

$$1 - \frac{\chi_M^2 - df_M}{\chi_B^2 - df_B},$$

where χ_M^2 is the chi-square value from the proposed model, df_M is the degrees of freedom in the proposed model, χ_B^2 is the chi-square value from the baseline model, and

df_B is the degrees of freedom in the baseline model. The CFI can range from zero to one; some sources suggest that values greater than 0.95 indicate good fit (Hu & Bentler, 1995), while others accept values greater than 0.90 (Bentler, 1990).

4.3 RESULTS

4.3.1 Identification of Eligible Cases

I began by determining the proportion of cases with complete data for each series of items. As shown in Table 4.1, the proportion of cases with complete data ranged from 82.1% to 93.1%. There were no significant differences between the incentive and control groups in the proportion of cases with complete data.

Table 4.1. Proportion of Cases with Complete Data, by Incentive Condition

	All cases	**Incentive**	**Control**	**Chi-square**
Patriotism (n=898)	82.1	82.6	81.4	0.23 (*p*=0.629)
Conscientiousness (n=898)	93.1	93.7	92.3	0.68 (*p*=0.408)
Impression Management (n=1016)	90.9	92.0	90.1	1.01 (*p*=0.315)
Consumer Sentiment				
$10 vs. 0 (n=529)	84.7	83.2	86.6	1.21 (*p*=0.271)
$5 vs. 0 (n=514)	85.2	84.0	86.6	0.68 (*p*=0.409)
$10 vs. $5 (n=579)	83.6	83.2	84.0	0.08 (*p*=0.776)

4.3.2 Comparison of Mean Responses

Table 4.2 shows the mean value for each series of items for the incentive and control groups. There were no significant differences in the mean overall scores for any of the series.

Table 4.2. Mean Index Score, by Incentive Condition

	Incentive	**Control**	***t*-test**
Patriotism (10 items)	31.8 (n=433)	31.9 (n=304)	-0.56 (*p*=0.577)
Conscientiousness (10 items)	42.2 (n=491)	42.2 (n=345)	-0.01 (*p*=0.994)
Impression Management (10 items)	28.4 (n=412)	28.5 (n=512)	-0.33 (*p*=0.742)
Consumer Sentiment (5 items)			
$10 vs. 0	18.0 (n=247)	17.7 (n=201)	0.71 (*p*=0.477)
$5 vs. 0	18.5 (n=237)	17.7 (n=201)	1.62 (*p*=0.105)
$10 vs. $5	18.0 (n=247)	18.5 (n=237)	-0.92 (*p*=0.359)

4.3.3 Internal Consistency Reliability

Table 4.3 shows the raw alpha coefficients for each series of items for the incentive and control groups. Generally, values of 0.70 or higher are considered acceptable. The alpha values ranged from 0.60 to 0.81, with the values for the impression management and consumer sentiment series falling just short of the 0.70 threshold. Generally, there were not significant differences between the alpha values for the incentive and control groups. However, for the patriotism series, the incentive group value was significantly lower than the control group value, suggesting that the incentive group's responses did not measure a single construct as well as the control group's responses did. Full correlation matrices are provided in Appendix F.

Table 4.3. Cronbach's Alpha, by Incentive Condition

	Incentive alpha	Control alpha	*W*-statistic
Patriotism	0.70 (n=433)	0.81 (n=304)	1.57 (p<0.001)
Conscientiousness	0.77 (n=491)	0.77 (n=345)	1.02 (p=0.423)
Impression Management	0.68 (n=412)	0.66 (n=512)	1.04 (p=0.339)
Consumer Sentiment			
$10 vs. 0	0.66 (n=247)	0.64 (n=201)	1.06 (p=0.331)
$5 vs. 0	0.60 (n=237)	0.64 (n=201)	1.11 (p=0.223)
$10 vs. $5	0.66 (n=247)	0.60 (n=237)	1.17 (p=0.112)

4.3.4 Differential Item Functioning

Tables 4.4 through 4.9 show the results of the differential item functioning (DIF) analysis. The first column of the tables lists each of the survey items included in the series. The second column shows the chi-square values resulting from ordinal logistic regression models predicting the response to each question where the independent variable was total score across the items in the series. The third column shows the chi-square values resulting from ordinal logistic regression models predicting response to each survey item where the independent variables were incentive receipt, total score across the items in the series, and an interaction term. The fourth column shows the difference between the chi-square values presented in the second and third columns. The final column shows the difference between the pseudo R^2 values produced by the two models; values less than 0.035 indicated negligible DIF, while values between 0.035 and 0.070 indicated moderate DIF, and values greater than or equal to 0.070 indicated large DIF (Gelin & Zumbo, 2003).

Significant $\chi^2(2)$ values in the fourth column suggested the presence of DIF, or that respondents in the incentive and control groups with identical amounts of the latent trait being measured would not provide the same response to that survey item. Two of the thirty-five comparisons shown were significant at the 0.05 level: "doing just enough work to get by" in the conscientiousness series from the Practicum telephone survey and "taken leave from work or school even though not really sick" in the impression management series from the mail survey. However, due to the large number of comparisons, Zumbo and Thomas (1997) suggest using a more stringent criteria for statistical significance – that *p* should be less than 0.01. Using this criterion, none of the items displayed differential item functioning. I also estimated the magnitude of DIF for each item by comparing the R^2 values produced by the two models; all of the resulting values were less than 0.035, again suggesting negligible DIF.

Table 4.4. Differential Item Functioning: Patriotism (n=737)

	$\chi^2(1)$ **from Model 1**	$\chi^2(3)$ **from Model 2**	$\chi^2(2)$ **: Difference between Model 1 and Model 2**	***Pseudo* R^2: Difference between Model 1 and Model 2**
Way democracy works	395.87 (p<0.001)	400.33 (p<0.001)	4.46 (p=0.108)	0.003
Political influence in the world	325.07 (p<0.001)	327.60 (p<0.001)	2.53 (p=0.282)	0.002
Economic achievements	395.20 (p<0.001)	396.23 (p<0.001)	1.03 (p=0.598)	0.001
Social security system	168.50 (p<0.001)	171.89 (p<0.001)	3.39 (p=0.184)	0.004
Scientific / technological achievements	248.25 (p<0.001)	251.20 (p<0.001)	2.95 (p=0.229)	0.003
Achievements in sports	295.78 (p<0.001)	297.61 (p<0.001)	1.83 (p=0.401)	0.002
Achievements in arts and literature	237.73 (p<0.001)	241.56 (p<0.001)	3.83 (p=0.147)	0.004
Armed forces	123.08 (p<0.001)	125.21 (p<0.001)	2.13 (p=0.345)	0.002
History	276.38 (p<0.001)	276.84 (p<0.001)	0.46 (p=0.795)	0.000
Fair and equal treatment of all groups	274.53 (p<0.001)	274.67 (p<0.001)	0.14 (p=0.932)	0.000

Table 4.5. Differential Item Functioning: Conscientiousness (n=836)

	$\chi^2(1)$ **from Model 1**	$\chi^2(3)$ **from Model 2**	$\chi^2(2)$ **: Difference between Model 1 and Model 2**	***Pseudo* R^2: Difference between Model 1 and Model 2**
Always prepared	274.13 (p<0.001)	274.54 (p<0.001)	0.41 (p=0.815)	0.000
Carry out plans	280.42 (p<0.001)	282.70 (p<0.001)	2.28 (p=0.320)	0.002
Pay attention to details	172.20 (p<0.001)	175.75 (p<0.001)	3.55 (p=0.169)	0.003
Waste time	363.46 (p<0.001)	364.03 (p<0.001)	0.57 (p=0.752)	0.000
Just enough work to get by	360.87 (p<0.001)	369.65 (p<0.001)	8.78* (p=0.012)	0.007
Don't see things through	343.52 (p<0.001)	345.87 (p<0.001)	2.35 (p=0.309)	0.002
Make plans and stick to them	304.18 (p<0.001)	305.90 (p<0.001)	1.72 (p=0.423)	0.001
Difficulty getting started doing work	430.89 (p<0.001)	433.60 (p<0.001)	2.71 (p=0.258)	0.002
Avoid duties	246.06 (p<0.001)	248.13 (p<0.001)	2.07 (p=0.355)	0.002
Get chores done right away	370.91 (p<0.001)	373.30 (p<0.001)	2.39 (p=0.303)	0.002

Table 4.6. Differential Item Functioning: Impression Management (n=924)

	$\chi^2(1)$ **from Model 1**	$\chi^2(3)$ **from Model 2**	$\chi^2(2)$ **: Difference between Model 1 and Model 2**	***Pseudo* R^2: Difference between Model 1 and Model 2**
Sometimes drive faster than speed limit	202.24 (p<0.001)	208.08 (p<0.001)	5.84 (p=0.054)	0.005
Never cover up mistakes	142.19 (p<0.001)	145.36 (p<0.001)	3.17 (p=0.205)	0.003
Sometimes tell lies if have to	373.16 (p<0.001)	377.11 (p<0.001)	3.95 (p=0.139)	0.003
Never swear	316.70 (p<0.001)	317.74 (p<0.001)	1.04 (p=0.595)	0.001
Never dropped litter on street	292.33 (p<0.001)	294.52 (p<0.001)	2.19 (p=0.335)	0.002
Occasions where taken advantage of someone	415.97 (p<0.001)	417.29 (p<0.001)	1.32 (p=0.517)	0.001
Taken leave from work/school even though not really sick	361.67 (p<0.001)	369.88 (p<0.001)	8.21* (p=0.016)	0.006
Never take things that don't belong to me	163.30 (p<0.001)	164.99 (p<0.001)	1.69 (p=0.430)	0.002
Said something bad about a friend behind his/her back	395.74 (p<0.001)	395.96 (p<0.001)	0.22 (p=0.896)	0.000
When hear people talking privately, avoid listening	264.93 (p<0.001)	265.99 (p<0.001)	1.06 (p=0.589)	0.001

Table 4.7. Differential Item Functioning: Consumer Sentiment: $10 vs. 0 (n=448)

	$\chi^2(1)$ **from Model 1**	$\chi^2(3)$ **from Model 2**	$\chi^2(2)$ **: Difference between Model 1 and Model 2**	***Pseudo* R^2: Difference between Model 1 and Model 2**
Personal financial situation compared to one year ago	210.53 (p<0.001)	212.98 (p<0.001)	2.45 (p=0.294)	0.003
Personal financial situation compared to one year from now	145.28 (p<0.001)	147.71 (p<0.001)	2.43 (p=0.297)	0.004
Business conditions in the next 12 months	345.27 (p<0.001)	347.88 (p<0.001)	2.61 (p=0.271)	0.000
Good/bad times for the country in next five years	350.49 (p<0.001)	351.08 (p<0.001)	0.59 (p=0.745)	0.003
Good/bad time to buy major household items	87.77 (p<0.001)	90.29 (p<0.001)	2.52 (p=0.284)	0.005

Table 4.8. Differential Item Functioning: Consumer Sentiment: $5 vs. 0 (n=438)

	$\chi^2(1)$ **from Model 1**	$\chi^2(3)$ **from Model 2**	$\chi^2(2)$ **: Difference between Model 1 and Model 2**	***Pseudo* R^2: Difference between Model 1 and Model 2**
Personal financial situation compared to one year ago	193.80 (p<0.001)	198.07 (p<0.001)	4.17 (p=0.124)	0.006
Personal financial situation compared to one year from now	120.23 (p<0.001)	120.37 (p<0.001)	0.14 (p=0.932)	0.000
Business conditions in the next 12 months	352.67 (p<0.001)	355.34 (p<0.001)	2.67 (p=0.263)	0.003
Good/bad times for the country in next five years	318.96 (p<0.001)	319.33 (p<0.001)	0.37 (p=0.831)	0.000
Good/bad time to buy major household items	78.91 (p<0.001)	82.51 (p<0.001)	3.60 (p=0.165)	0.007

Table 4.9. Differential Item Functioning: Consumer Sentiment: $10 vs. $5 (n=484)

	$\chi^2(1)$ **from Model 1**	$\chi^2(3)$ **from Model 2**	$\chi^2(2)$ **: Difference between Model 1 and Model 2**	***Pseudo* R^2: Difference between Model 1 and Model 2**
Personal financial situation compared to one year ago	165.34 (p<0.001)	165.87 (p<0.001)	0.53 (p=0.767)	0.001
Personal financial situation compared to one year from now	171.23 (p<0.001)	172.91 (p<0.001)	1.68 (p=0.432)	0.002
Business conditions in the next 12 months	383.94 (p<0.001)	387.59 (p<0.001)	3.65 (p=0.161)	0.003
Good/bad times for the country in next five years	341.86 (p<0.001)	346.77 (p<0.001)	4.91 (p=0.086)	0.005
Good/bad time to buy major household items	120.08 (p<0.001)	120.72 (p<0.001)	0.64 (p=0.726)	0.001

4.3.5 Confirmatory Factor Analysis

I estimated a series of factor models that tested for increasing levels of invariance between the incentive and control group responses. In the first step, I determined whether the proposed measurement models fit the data acceptably well for both the incentive and control groups. To assess this, I estimated a one-factor model separately for the incentive and control groups for each set of items. These models were calculated based on covariance matrices, which are provided in Appendix F.

The fit indices for each model are shown in Table 4.10. As discussed earlier, when sample sizes are large, the chi-square statistic is overly sensitive; as a result, though I still present the (highly significant) chi-square values here, I focused on the other fit indices to determine whether or not the model fit the data acceptably well. Most of the other fit indices suggested good fit (RMSEA<0.07, SRMR<0.05, CFI<0.95) or were very close to meeting the cutoff for good fit. In the few exceptions, the indices still generally

met, or came close to, the cutoffs for fair fit (SRMR<0.08, CFI<0.90). For example, though several of the SRMR values were above 0.05, none of them were greater than 0.08. Overall, these results suggested that the models fit the data sufficiently well that they provide a reasonable basis to test for measurement invariance across the two groups.

Table 4.10. Fit Indices for Initial Measurement Models, by Incentive Condition

	Chi-square (df, *p*-value)	**RMSEA (90% CI)**	**SRMR**	**CFI**
Patriotism				
Incentive	132.81 (df=35, $p<0.001$)	0.08 (0.07, 0.10)	0.06	0.88
Control	62.26 (df=35, $p<0.01$)	0.05 (0.03, 0.07)	0.04	0.98
Conscientiousness				
Incentive	196.08 (df=35, $p<0.01$)	0.11 (0.10, 0.13)	0.07	0.90
Control	155.29 (df=35, $p<0.01$)	0.11 (0.10, 0.13)	0.08	0.89
Impression Management				
Incentive	107.04 (df=35, $p<0.001$)	0.07 (0.06, 0.09)	0.06	0.90
Control	133.68 (df=35, $p<0.01$)	0.08 (0.07, 0.09)	0.06	0.88
Consumer Sentiment				
$10 Incentive	3.34 (df=5, $p=0.65$)	0.00 (0.00, 0.07)	0.03	1.00
$5 Incentive	4.55 (df=5, $p=0.47$)	0.00 (0.00, 0.09)	0.03	1.00
Control	6.58 (df=5, $p=0.27$)	0.04 (0.00, 0.11)	0.04	0.99

Next, I tested for configural invariance – this suggests that, for both groups, there is the same number of factors and that the same items load on each of these factors. To test for configural invariance, I estimated a single model for both the incentive and control groups in which all of the items in the series were required to load onto a single factor, but in which the factor loadings were permitted to differ for each group. Again, I

reviewed the fit indices produced by these models (Table 4.11). If the models fit the data acceptably well, this suggests that configural invariance exists. Again, the majority of the fit indices suggested good fit, and those that did not suggested at least fair fit. Thus, these results confirm that a single-factor model is appropriate for both groups for all of the sets of items.

Table 4.11. Fit Indices for Configural Invariance Models

	Chi-square (df, *p*-value)	**RMSEA (90% CI)**	**SRMR**	**CFI**
Patriotism	195.07 (df=70, $p<0.01$)	0.07 (0.06, 0.08)	0.04	0.94
Conscientiousness	351.37 (df=70, $p<0.01$)	0.11 (0.10, 0.12)	0.08	0.89
Impression Management	240.73 (df=70, $p<0.01$)	0.08 (0.07, 0.09)	0.06	0.89
Consumer Sentiment				
$10 Incentive vs. Control	9.92 (df=10, $p=0.46$)	0.00 (0.00, 0.07)	0.04	1.00
$5 Incentive vs. Control	11.40 (df=10, $p=0.33$)	0.02 (0.00, 0.08)	0.04	1.00
$10 Incentive vs. $5 Incentive	8.17 (df=10, $p=0.61$)	0.00 (0.00, 0.06)	0.03	1.00

In the next step, I tested for metric invariance – this suggests that the factor loadings are equivalent for the two groups. To test for metric invariance, I estimated a single model for both the incentive and control groups in which all of factor loadings were constrained to be equivalent. Again, I reviewed the fit indices produced by these models (Table 4.12); they generally suggested that the metric invariance models fit the data acceptably well.

Table 4.12. Fit Indices for Metric Invariance Models

	Chi-square (df, *p*-value)	**Diff Btwn Chi-square values**	**RMSEA (90% CI)**	**SRMR**	**CFI**
Patriotism	206.05 (df=79, $p<0.01$)	10.98 (df=9, $p=0.277$)	0.07 (0.06, 0.08)	0.03	0.94
Conscientiousness	360.90 (df=79, $p<0.01$)	9.53 (df=9, $p=0.390$)	0.11 (0.10, 0.12)	0.08	0.89
Impression Management	255.42 (df=79, $p<0.01$)	14.69 (df=9, $p=0.100$)	0.07 (0.06, 0.08)	0.07	0.89
Consumer Sentiment					
$10 Incentive vs. Control	21.08 (df=14, $p=0.12$)	11.16 (df=4, $p=0.025$)	0.05 (0.00, 0.09)	0.07	0.98
$5 Incentive vs. Control	29.34 (df=14, $p=0.01$)	17.94 (df=4, $p=0.001$)	0.07 (0.03, 0.10)	0.08	0.95
$10 Incentive vs. $5 Incentive	12.53 (df=14, $p=0.56$)	4.36 (df=4, $p=0.359$)	0.00 (0.00, 0.05)	0.04	1.00

I also compared the chi-square values produced by these models to the values produced by the configural invariance models. A significant difference between these values suggested that model fit was significantly reduced by adding this constraint and that metric invariance could not be established. The difference was not significant for the patriotism, conscientiousness, or impression management measurement models, again supporting metric invariance. However, for the consumer sentiment models, some of the differences were significant – namely for the models that compared each of the incentive groups to the control group. Thus, for this measurement model, the metric invariance assumption was not supported.

As a final step, I tested for partial invariance for the consumer sentiment model. Partial invariance suggests that invariance holds for some of the items but not for others. To determine which items were keeping the model from achieving full metric invariance, I reviewed the modification indices (MIs) produced by the metric invariance models. These values indicated the gain in the chi-square value that would be achieved by releasing the constraint that the factor loading for that item must be equivalent across the two groups. As shown in Table 4.13, for both models, the modification index for the first item indicated that there would be a significant gain in the chi-square value by releasing the constraint on this item's factor loading. This suggested that invariance did not hold for this particular item.

Table 4.13. Modification Indices Produced by Metric Invariance Models

	$10 vs. Control	**$5 vs. Control**
Personal financial situation compared to one year ago	9.72**	13.23**
Personal financial situation compared to one year from now	1.75	0.01
Business conditions in the next 12 months	0.71	3.37
Good/bad times for the country in next five years	0.01	0.05
Good/bad time to buy major household items	1.23	2.07

* $p<0.05$; ** $p<0.01$

As a result, I released the equality constraint for this item and re-estimated the metric invariance model to test for partial invariance. Table 4.14 shows the fit indices for these models. These models fit the data quite well, with RMSEA values less than 0.05, SRMR values less than 0.07, and CFI values greater than 0.95. Additionally, the differences between the metric invariance chi-square values and the configural invariance chi-square values were no longer significant.

Table 4.14. Fit Indices for Metric Invariance Models with One Equality Constraint Removed

	Chi-square (df, *p*-value)	**Diff Btwn Chi-square values**	**RMSEA (90% CI)**	**SRMR**	**CFI**
Consumer Sentiment					
$10 Incentive vs. Control	11.95 (df=13, p=0.53)	2.03 (df=3, p=0.566)	0.00 (0.00, 0.06)	0.05	1.00
$5 Incentive vs. Control	14.70 (df=13, p=0.33)	3.30 (df=3, p=0.348)	0.02 (0.00, 0.07)	0.05	0.99

As a final test, I again reviewed the modification indices produced by these models to ensure that none of them were statistically significant. As shown in Table 4.15, the values were quite small and none of them were statistically significant, suggesting that the factor loadings were equivalent for the remaining four items.

Table 4.15. Modification Indices After Releasing One Equality Constraint

	$10 vs. Control	**$5 vs. Control**
Personal financial situation compared to one year ago	--	--
Personal financial situation compared to one year from now	1.09	0.06
Business conditions in the next 12 months	0.01	0.78
Good/bad times for the country in next five years	0.81	1.38
Good/bad time to buy major household items	0.86	1.48

* $p<0.05$; ** $p<0.01$

Table 4.16 shows the unstandardized factor loadings for the final models. For example, in the incentive group, the unstandardized loading of 0.89 for the first item suggests that a one-point difference in the consumer sentiment factor score would be associated with a 0.89-point difference in the response to this item. Table 4.17 shows the standardized loadings; these are estimated correlations between the item and the factor. In both cases, for the item where the equality constraint was released, the factor loading was smaller in the incentive group than it was in the control group. This suggests that incentive group respondents' personal financial situation as compared to one year ago was less important in relation to their overall sentiment as consumers than it was for control group respondents.

Table 4.16. Unstandardized Factor Loadings: Consumer Sentiment

	$10 vs. Control		$5 vs. Control	
	Incentive	**Control**	**Incentive**	**Control**
Personal financial situation compared to one year ago	0.89**	3.48**	1.03**	3.14**
Personal financial situation compared to one year from now	1.90**		1.41**	
Business conditions in the next 12 months	4.60**		4.40**	
Good/bad times for the country in next five years	4.39**		3.87**	
Good/bad time to buy major household items	1.00[a]		1.00[a]	

a. Fixed at 1.00 to scale the factor. Not tested for statistical significance.

* $p<0.05$; ** $p<0.01$

Table 4.17. Standardized Factor Loadings: Consumer Sentiment

	$10 vs. Control		$5 vs. Control	
	Incentive	**Control**	**Incentive**	**Control**
Personal financial situation compared to one year ago	0.31**	0.63**	0.20**	0.62**
Personal financial situation compared to one year from now	0.44**		0.37**	
Business conditions in the next 12 months	0.77**		0.79**	
Good/bad times for the country in next five years	0.76**		0.71**	
Good/bad time to buy major household items	0.21[a]		0.23[a]	

a. Not tested for statistical significance.

* $p<0.05$; ** $p<0.01$

4.4 DISCUSSION

Prepaid cash incentives appear to be an effective tool for increasing survey response rates; however, researchers have expressed concern that their use also might have unintended effects on survey estimates. To date, existing literature has found little support for this concern. Most of these studies have compared the response distributions of incentive and control group respondents – but there are several drawbacks to this approach. For sets of items that are intended to measure a latent construct, there is an alternative analytic approach that provides a more rigorous test of differences across groups – testing for measurement invariance in responses provided by incentive and control group respondents. These methods assess how well the observed responses to the items measure the latent construct that they are intended to measure. Measurement invariance exists if respondents with the same degree of the latent construct have the same probability of endorsing survey items.

In this chapter, I utilized several methods to test for measurement invariance in four series of items drawn from recent large scale surveys that included incentive experiments. First, I calculated alpha values for each experimental group for each series of items. For one of the four series, there was a significant difference in the alpha values observed for the incentive and control groups; for the patriotism series from the Practicum telephone survey, the incentive group value was significantly lower than the control group value, suggesting that the incentive group's responses did not measure a single construct as well as the control group's responses did.

I also utilized ordinal logistic regression models to assess the prevalence of differential item functioning (DIF); DIF did not appear to be present for any of the items in any of the four series. Finally, I used confirmatory factor analysis to test for configural

and metric invariance. All four of the series exhibited configural invariance. However, the consumer sentiment series did not exhibit metric invariance, suggesting that the factor loadings could not be assumed to be equivalent in the incentive and control groups for all five items in the series. Reviewing the modification indices suggested that the equality constraint should be removed for the item asking respondents to compare their current financial situation to their financial situation one year earlier. Removing this constraint allowed the model to achieve partial invariance.

The resulting model suggested that the factor loading for this item was smaller in the two groups that received an incentive as compared to the control group. This suggests that the comparison of their current financial situation to that of one year prior was not as predictive of incentive group respondents' consumer sentiment as it was for control group respondents. This raises the possibility that paying respondents, even token amounts, may, at least temporarily, affect their assessment of their financial situation and its relationship to their expectations for future spending behavior. For example, although these differences did not reach statistical significance, respondents in both the $10 and $5 incentive groups were more likely than control group respondents to feel that their current financial situation was better than that of one year prior (46.4%, 44.1%, and 39.8%, respectively).

As discussed in earlier chapters, there are limitations to the datasets utilized in this chapter. For example, the Practicum and mail surveys both utilized somewhat specialized populations (heads of households with listed phone numbers and registered voters, respectively) that may react to incentives differently than would the general population. Furthermore, for the Survey of Consumers, I was not able to determine which cases had received refusal conversion incentives. However, overall, these results largely suggest that

measurement invariance was established between the responses of incentive and control group respondents. This is reassuring, as it suggests that prior findings of no difference between incentive and control group responses are likely, in fact, valid.

CHAPTER 5

SUMMARY AND CONCLUSIONS

Extensive research indicates that prepaid incentives are an effective tool for increasing survey response rates. As nonresponse continues to rise, the use of incentives in survey research also appears to be increasing. It is important to consider the effect that prepaid incentives might have on various sources of error. As prepaid incentives clearly convince some sample members to participate who otherwise would not have done so, it is reasonable to believe that they also may influence the way that respondents act *during* survey interviews. The use of incentives conceivably could either increase or decrease the magnitude of measurement error in survey estimates. This would be useful information for practitioners to have when considering the use of incentives. Nevertheless, the majority of incentives research has focused narrowly on the effect that incentives have on response rates.

Existing literature tends to take one of three approaches for investigating the effect of incentives on measurement error: (1) comparing survey responses provided by incentive and control group respondents, (2) comparing the relative accuracy of survey responses as compared to validation data, and (3) comparing (lack of) effort indicators. Typically, these studies have concluded that incentives either result in small increases in quality or have no effect.

However, these studies are limited in both number and scope. For example, studies comparing effort indicators tend to focus narrowly on item nonresponse and responses to open-ended items. Additionally, research comparing response distributions typically assumes that incentives affect all types of survey items in the same manner. Furthermore, such studies tend to examine the effect of the incentive on each item in

isolation, without considering the possible impact on relationships between variables. Finally, existing literature does not typically consider the possibility that the magnitude of the incentive effect might differ according to respondent characteristics.

This dissertation included three studies aimed at addressing these limitations. The first study evaluated the effect of a prepaid incentive on respondent effort, the second investigated the impact of a prepaid incentive on responses to sensitive items, and the third tested for measurement invariance in responses provided by incentive and control group respondents. The key results of these studies are provided in Table 5.1.

The first study, presented in Chapter 2, assessed the effect of a $5 prepaid cash incentive on respondent effort in a national telephone survey. The incentive led to a significant increase in the response rate (from 11% to 23%), contact rate (from 55% to 58%), and cooperation rate (from 20% to 39%), as well as a significant reduction in the refusal rate (from 41% to 32%). It resulted in approximately a ten percent reduction in the cost per complete (from $64 to $58). There were no significant differences in the demographic characteristics of the incentive and control group respondents.

The study included twelve effort indicators. The effect of the incentive on each of these is noted in Table 5.1. The incentive led to significant reductions in two of the twelve indicators. First, respondents who received an incentive skipped a significantly smaller proportion of the items (2% vs. 3%) and were significantly less likely to skip at least one item (54% vs. 61%). Separately reviewing the effect on "don't know" and "refuse" responses indicated that the significance of this effect was limited to "don't know" responses. Second, incentive group respondents spent significantly less time per question that they were asked (from 17 seconds to 16 seconds). However, the incentive

did not have a significant impact on the other ten indicators. It also did not have a significant effect on a satisficing index which measured the proportion of items for which respondents displayed at least one cognitive shortcut.

Table 5.1. Key Findings

Chapter 1: Respondent Effort in a National Telephone Survey with $5 Incentive (n=900)	
Indicator	***Effect of Incentive***
Item nonresponse	Significant reduction (from 3% to 2 %)
Responses to open-ended items	None
Straight-lining / non-differentiation	None
Acquiescence	None
Response order effects	None
Lack of attention to important exclusions	None
Round values	None
Estimation strategy for numerical responses	None
Underreporting for filter items	None
Interviewer length	Significant reduction (from 17 seconds/question to 16 seconds/question)
Accuracy as compared to frame data	None
Interviewer reports of effort	None
Satisficing index	None
Chapter 2: Responses to Sensitive Questions in a Mail Survey of Registered Voters with $5 Incentive (n=1016)	
Indicator	***Effect of Incentive***
Impression management scale	None
Reports of undesirable behaviors and attitudes High sensitivity Medium sensitivity Low sensitivity	 Significant increase (from 25% to 27%) None None
Item nonresponse High sensitivity Medium sensitivity Low sensitivity	 None None None
Nonresponse bias: three voting items	Significant reduction for one item (2004)
Measurement bias: three voting items	None
Chapter 3: Measurement Invariance in Three Recent Surveys	
Indicator	***Effect of Incentive***
Internal consistency reliability: four scales	Significant reduction for one of four scales (patriotism)
Differential item functioning: four scales	None
Confirmatory factor analysis: four scales	Only partial metric invariance for one of four scales (consumer sentiment)

I hypothesized that the magnitude of the incentive effect might differ according to item characteristics, such as position in the questionnaire. For the two indicators where the incentive had a significant overall effect on effort (item nonresponse and interview length), I found that the significance of the incentive effect was limited to the second half of the interview. For the other three indicators where there were sufficient items in both halves of the questionnaire to test this hypothesis (acquiescence, response order effects, and round values), the incentive did not have a significant effect on effort in either half of the interview. I also hypothesized that the incentive might have a greater effect on responses to attitude questions than it did on responses to factual items; for the one indicator where there were sufficient items to test this hypothesis (acquiescence), I did not observe a significant effect of the incentive on either type of items.

Finally, I investigated the possibility that the magnitude of the incentive's effect on effort might vary depending on respondent characteristics. I hypothesized that the incentive would have a greater effect on respondents with greater cognitive ability than on respondents with less cognitive ability, and that it would have a greater effect on less conscientious respondents than on more conscientious respondents; I found little support for either of these hypotheses. I also predicted that the effect of the incentive would be limited to those respondents who recalled receiving it. While there was minimal evidence of improved effort among respondents who recalled the incentive, those who did not recall receiving the incentive appeared to provide lower quality data than the control group respondents; they were significantly more likely to take shortcuts for four of the ten indicators that could be measured at the individual respondent level.

Chapter 3 evaluated the effect that a prepaid cash incentive had on self-presentation concerns and responses to sensitive questions in a mail survey of registered voters. The incentive led to a significant increase in the response rate (from 20% to 41%) and about a 10% reduction in the cost per complete (from $32 to $30). It also resulted in improved representation of younger people.

The incentive led to a significant increase in reports of highly undesirable attitudes and behaviors (from 25% to 27%) and had no effect on responses to less sensitive items; this suggests that item sensitivity may play a role in the magnitude of the incentive effect. However, the incentive had little influence on the extent of item nonresponse (approximately 4% of items skipped in both groups) – regardless of item sensitivity – or on impression management concerns (mean response of 2.8 in both groups, on a scale of 1 to 4). Respondent demographics, such as age and race, had no impact on the magnitude of the incentive effect for these outcomes.

Voting records were available for three recent elections for all sample members; the survey also included three items asking respondents whether or not they had voted in these elections. The survey considerably overestimated voting behavior, on the magnitude of 20-30 percentage points for each election. There were significant nonresponse and measurement biases for all three items. The incentive resulted in a general pattern of reduced nonresponse bias (by approximately 5 percentage points) and increased measurement bias (by 1 to 4 percentage points); however, these effects generally were not significant. I also hypothesized that the incentive would have a greater effect on nonvoters than on voters – for both participation in the survey and accuracy of voting reports. The incentive did appear to have a significantly greater effect on the

participation decision for nonvoters than for voters (particularly consistent voters), but the effect of the incentive on accuracy did not vary significantly by voting history.

Finally, in Chapter 4, I tested for measurement invariance between the responses provided by incentive and control group respondents. This analysis utilized four sets of items intended to measure latent constructs that came from three different surveys that included prepaid incentive experiments; in all three studies the use of a prepaid incentive significantly increased the response rate.

I evaluated internal consistency reliability by calculating Cronbach's alpha for each scale for each experimental group. Generally the alpha values for the incentive and control groups were not significantly different from each other; however, for one scale – the patriotism items from the Practicum survey – the alpha value was significantly lower in the incentive group than in the control group (0.70 and 0.81, respectively). This suggests that the incentive group respondents' answers to these items were not as consistent as were those of the control group.

I also utilized two approaches to test for measurement invariance: differential item functioning and confirmatory factory analysis. There was no evidence that the incentive led to differential item functioning for any of the items. However, the confirmatory factor analysis approach demonstrated that full metric invariance could not be established for the consumer sentiment items from the Survey of Consumers, suggesting that the factor loadings for at least some of these items were not equivalent in the incentive and control groups. In particular, metric invariance did not hold for an item that asked respondents to rate their personal financial situation as compared to one year ago. When an incentive was offered, it appeared that this item was less important to

respondents' determination of their overall sentiment as consumers than it was for control group respondents. This raises the possibility that paying respondents, even token amounts (such as $5 or $10), may temporarily affect their assessment of their financial situation and their expectations for future spending behavior. However, overall these results largely suggest that measurement invariance was established.

In the three studies conducted as part of this dissertation, prepaid incentives had minimal effect on measurement error. There was some evidence that *item* characteristics, such as sensitivity or position in the questionnaire, may affect the magnitude of the incentive effect. Additionally, the possibility that items referring to respondents' financial situation are sensitive to incentive effects may deserve further research. However, there was little to support the hypothesis that the *respondent* characteristics explored in this dissertation play a role in the magnitude of the effect. The unexpected finding that incentive group respondents who did not remember receiving the incentive provided lower quality data than control group respondents also may deserve further research.

There were limitations to this research that I hope will be addressed in future work. Although I aimed to determine the effect of prepaid incentives on measurement error, I generally was restricted to analysis of indirect indicators of error. For example, increases in effort may be associated with reduced error, but we cannot be certain that this always is the case. Future research should aim to compare survey reports to validation data whenever possible in order to more concretely assess the effect of incentives on measurement error.

Also, most of the incentive experiments utilized in this dissertation compared a prepaid $5 cash incentive to a control group; it would be useful to know whether the

effect of the incentive on data quality would vary according to incentive value or delivery timing. For example, it is possible that incentives of larger value could have slightly larger effects on measurement error; either by way of creating a greater sense of obligation or by recruiting more reluctant respondents to participate. Additionally, promised incentives could lead to improvements in response quality if respondents believe that a certain quality of response is required in order to qualify for the incentive. Finally, it seems that, by recruiting resistant sample members who have already refused to participate, refusal conversion incentives might lead to increases in measurement error; the effect of such incentives on data quality deserves further research.

Additionally, some of the surveys, especially the telephone survey utilized in Chapter 2, had rather low response rates. This suggests that these studies look at the effect of incentives among those who are most willing to participate; as such individuals may be inclined to participate in surveys even without an incentive, incentives may have a smaller effect on the behavior of these people than they would have on more reluctant respondents. It would be useful to know whether the results presented here would be replicated in surveys that are able to achieve higher response rates.

Finally, several of the studies focused on somewhat specialized populations (e.g., heads of households with listed telephone numbers, registered voters living in Maryland); it is possible that the results observed here would not be replicated in general population surveys. For example, having a listed phone number or registering to vote may suggest a greater openness to being contacted for things like survey research; again, incentives may have less of an effect on such individuals than they would on more reluctant respondents who have unlisted phone numbers.

In sum, prepaid incentives seemed to have had limited effect on measurement error. There also were several other positive outcomes associated with the use of incentives, such as a significant increase in the response rate, improved advance letter recall, a reduction in the cost per complete, and increased representation of youth. Moving forward, these results should be reassuring for researchers considering the use of prepaid incentives.

APPENDIX A

SURVEY MATERIALS, JPSM PRACTICUM SURVEY

A1. ADVANCE LETTER

[DATE]

Dear [NAME],

Researchers at The University of Maryland are conducting an important nationwide study about Americans' health and their views on various social issues. We are asking a scientific random sample of individuals to take part in a short interview. A few days from now you will receive a phone call from Princeton Data Source. If the call comes at an inconvenient time, the interviewer will be happy to set an appointment to call back at a better time.

Your help is voluntary but very important. The answers you give will be confidential, and we will take all possible steps to protect your privacy. Your answers will be used for research only.

[INCENTIVE: We have included a token of our appreciation for your participation. / CONTROL: We thank you in advance for your participation.] Your assistance is crucial to the success of this research.

Sincerely,

Dr. Stanley Presser

University of Maryland

A2. QUESTIONNAIRE

MAIN INTRODUCTION
Hello, my name is _____, and I am calling on behalf of the University of Maryland. May I please speak with [RESPONDENT NAME]?

ONCE TARGET RESPONDENT IS ON THE PHONE:
[IF RESPONDENT DID NOT ANSWER PHONE, REPEAT: Hello, my name is _____, and I am calling on behalf of the University of Maryland.] We're conducting a nationwide study on health, economic and other issues and we would like to include your opinions. Your participation is voluntary, and your responses will be used for research purposes only. To begin...
[READ IF NECESSARY: The interview will only take about 20 minutes to complete.]

VOICEMAIL MESSAGE (LEAVE ONLY ONCE -- THE FIRST TIME A CALL GOES TO VOICEMAIL):
Hello, I am calling on behalf of the University of Maryland. We're conducting a nationwide study on health, economic and other issues. This is NOT a sales call. I am sorry we missed you today and will try to reach you again. If you would like, please call us back at 1-800-887-3150 Monday through Friday 9 AM- 11:00 PM Eastern Daylight Time or 10:00 AM -10:00PM Eastern Daylight Time on Saturday and Sunday to schedule the interview. Have a good (day/evening). [SCHEDULE CALLBACK]

ASK ALL
SEX Respondent's sex (DO NOT ASK; RECORD BY OBSERVATION) {Formerly Q65}
1 Male
2 Female

START TIMING MODULE 1

SECTION 1: GENERAL HEALTH
ASK RANDOM HALF SAMPLE (EXPERIMENT1=1)
Q1A Would you say your health in general is excellent, very good, good, fair, or poor?
1 Excellent
2 Very good
3 Good
4 Fair
5 Poor
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE (EXPERIMENT1=2)

Q1B Would you say your health in general is poor, fair, good, very good, or excellent?

1 Poor
2 Fair
3 Good
4 Very good
5 Excellent
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF DESCRIBED HEALTH STATUS (Q1A=1-5 or Q1B=1-5)

Q2 Why do you feel that your health is [INSERT RESPONSE FROM Q1a or Q1b]? [OPEN-END; RECORD VERBATIM RESPONSE]

1 [RECORD VERBATIM RESPONSE]
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE (EXPERIMENT2=1)

Q3A People do different things in order to stay healthy. Which of the following do you think is the MOST important thing for a person to do in order to stay healthy – eat right, get enough sleep, reduce stress, have a yearly physical, or get regular exercise?[READ IF NECESSARY: If you had to choose just one, which do you think is most important?]

1 Eat right
2 Get enough sleep
3 Reduce stress
4 Have a yearly physical
5 Get regular exercise
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE (EXPERIMENT2=2)

Q3B People do different things in order to stay healthy. Which of the following do you think is the MOST important thing for a person to do in order to stay healthy – get regular exercise, have a yearly physical, reduce stress, get enough sleep, or eat right? [READ IF NECESSARY: If you had to choose just one, which do you think is most important?]

1 Get regular exercise
2 Have a yearly physical
3 Reduce stress
4 Get enough sleep
5 Eat right
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

[READ TO ALL:] The next questions are about foods you may have eaten in the past 7 days.

[RANDOMIZE Q4A1/Q4B1 WITH Q5A1/Q5B1; RANDOM HALF WILL GET Q4A1/Q4B1 FIRST (EXPERIMENT16=1) AND RANDOM HALF WILL GET Q5A1/Q5B1 FIRST (EXPERIMENT16=2)]

ASK RANDOM HALF SAMPLE (EXPERIMENT3=1)

Q4A1 In the past 7 days, how many servings of fruit did you eat? [READ IF NECESSARY: A serving of fruit is equal to about one half cup of fruit.]
[IF PROVIDES "PER DAY" RESPONSE, ASK: So how many servings of fruit would that be in the past 7 days?]

___ [RECORD EXACT NUMBER 0-97]
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE (EXPERIMENT3=2)

Q4B1 In the past 7 days, how many servings of fruit did you eat? Please do not include apples, bananas or oranges. [READ IF NECESSARY: A serving of fruit is equal to about one half cup of fruit.]
[IF PROVIDES "PER DAY" RESPONSE, ASK: So how many servings of fruit would that be in the past 7 days?]

___ [RECORD EXACT NUMBER 0-97]
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

INT_4 INTERVIEWER: DO NOT ASK; CODE RESPONDENT'S APPROACH TO ANSWERING QUESTION: IF R DID NOT SAY HOW, RECORD AS "DON'T KNOW"; RECORD AS MANY AS APPLY

1 Counted each individual serving (e.g., 1 serving Tuesday plus 3 servings Friday)
2 Used average daily servings to arrive at answer (e.g., I had 2 per day)
3 Thought about types of fruits and added them up
4 Other (SPECIFY)
98 (DO NOT READ) Don't know

ASK RANDOM HALF SAMPLE (EXPERIMENT4=1)

Q5A1 In the past 7 days, how many servings of vegetables did you eat? [READ IF NECESSARY: A serving of vegetables is equal to about one half cup of vegetables.]
[IF PROVIDES "PER DAY" RESPONSE, ASK: So how many servings of vegetables would that be in the past 7 days?]

___ [RECORD EXACT NUMBER 0-97]
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE (EXPERIMENT4=2)

Q5B1 In the past 7 days, how many servings of vegetables did you eat? Please do not include carrots, beans, or lettuce. [READ IF NECESSARY: A serving of vegetables is equal to about one half cup of vegetables.]
[IF PROVIDES "PER DAY" RESPONSE, ASK: So how many servings of vegetables would that be in the past 7 days?]

___ [RECORD EXACT NUMBER 0-97]
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

INT_5 INTERVIEWER: DO NOT ASK; CODE RESPONDENT'S APPROACH TO ANSWERING QUESTION: IF R DID NOT SAY HOW, RECORD AS "DON'T KNOW"; RECORD AS MANY AS APPLY

1 Counted each individual serving (e.g., 2 servings Tuesday plus 3 on Friday)
2 Used average daily servings to arrive at answer (e.g., I had 2 per day)
3 Thought about types of vegetables and added them up
4 Other (SPECIFY)
98 (DO NOT READ) Don't know

[RANDOMIZE ORDER OF Q6A1 / B1 / C1 / D1 / E1 / F1]
[READ TO ALL:] Has a doctor or other health professional EVER told you that you have any of the following?

ASK ALL

Q6A1 [IF RANDOMIZED 2ND-THRU-LAST, INSERT: How about] Diabetes or sugar diabetes [IF FEMALE (SEX=2), INSERT: other than during pregnancy]? [READ IF NECESSARY: Has a doctor or other health professional EVER told you that you have this condition?]
[INTERVIEWER NOTE: Do not accept self-diagnosed or diagnosed by a person who is not a doctor or other health professional]

1 Yes
2 No
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF DIAGNOSED WITH DIABETES (Q6A1=1)

Q6A2 How old were you when you were first diagnosed with diabetes or sugar diabetes? [READ IF NECESSARY: Just your best guess is fine.]

___ years old [RECORD EXACT AGE 1-96]
0 Less than 1 year old
97 97 years old or older
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q6B1 [IF RANDOMIZED 2ND-THRU-LAST, INSERT: How about] Hypertension or high blood pressure? [READ IF NECESSARY: Has a doctor or other health professional EVER told you that you have this condition?]
[INTERVIEWER NOTE: Do not accept self-diagnosed or diagnosed by a person who is not a doctor or other health professional]

1 Yes
2 No
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF DIAGNOSED WITH HYPERTENSION OR HIGH BLOOD PRESSURE (Q6B1=1)

Q6B2 How old were you when you were first diagnosed with hypertension or high blood pressure?
[READ IF NECESSARY: Just your best guess is fine.]

___ years old [RECORD EXACT AGE 1-96]
0 Less than 1 year old
97 97 years old or older
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q6C1 [IF RANDOMIZED 2ND-THRU-LAST, INSERT: How about] Asthma? [READ IF NECESSARY: Has a doctor or other health professional EVER told you that you have this condition?]
[INTERVIEWER NOTE: Do not accept self-diagnosed or diagnosed by a person who is not a doctor or other health professional]

1 Yes
2 No
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF DIAGNOSED WITH ASTHMA (Q6C1=1)

Q6C2 How old were you when you were first diagnosed with asthma?
[READ IF NECESSARY: Just your best guess is fine.]

___ years old [RECORD EXACT AGE 1-96]
0 Less than 1 year old
97 97 years old or older
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q6D1 [IF RANDOMIZED 2ND-THRU-LAST, INSERT: How about] Arthritis? [READ IF NECESSARY: Has a doctor or other health professional EVER told you that you have this condition?]
[INTERVIEWER NOTE: Do not accept self-diagnosed or diagnosed by a person who is not a doctor or other health professional]

1 Yes
2 No
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF DIAGNOSED WITH ARTHRITIS (Q6D1=1)

Q6D2 How old were you when you were first diagnosed with arthritis?
[READ IF NECESSARY: Just your best guess is fine.]

___ years old [RECORD EXACT AGE 1-96]
0 Less than 1 year old
97 97 years old or older
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q6E1 [IF RANDOMIZED 2ND-THRU-LAST, INSERT: How about] Heart disease? [READ IF NECESSARY: Has a doctor or other health professional EVER told you that you have this condition?]
[INTERVIEWER NOTE: Do not accept self-diagnosed or diagnosed by a person who is not a doctor or other health professional]

1 Yes
2 No
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF DIAGNOSED WITH HEART DISEASE (Q6E1=1)

Q6E2 How old were you when you were first diagnosed with heart disease?
[READ IF NECESSARY: Just your best guess is fine.]

___ years old [RECORD EXACT AGE 1-96]
0 Less than 1 year old
97 97 years old or older
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q6F1 [IF RANDOMIZED 2ND-THRU-LAST, INSERT: How about] Anemia? [READ IF NECESSARY: Has a doctor or other health professional EVER told you that you have this condition?]
[INTERVIEWER NOTE: Do not accept self-diagnosed or diagnosed by a person who is not a doctor or other health professional]

1 Yes
2 No
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF DIAGNOSED WITH ANEMIA (Q6F1=1)

Q6F2 How old were you when you were first diagnosed with anemia?
[READ IF NECESSARY: Just your best guess is fine.]

___ years old [RECORD EXACT AGE 1-96]
0 Less than 1 year old
97 97 years old or older
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q7 In 2010, were you a patient in a hospital overnight? Do not include an overnight stay in the emergency room.

1 Yes
2 No
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF HOSPITAL INPATIENT IN 2010 (Q7=1)

Q8 How many times were you a patient in a hospital overnight or longer during 2010? Do not count the total number of nights, just the total number of hospital admissions for stays which lasted 1 or more nights.

___ [RECORD EXACT NUMBER 1-97]
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q9A During 2010, how many times did you see a doctor or other health care professional about your health at a doctor's office, a clinic, hospital emergency room, at home or some other place? [IF HOSPITAL INPATIENT (Q7=1), READ: Do not include times you were hospitalized overnight.]
[READ IF NECESSARY: How many times would that be for all of 2010?] [IF DK or REF, READ: Just your best guess is fine.]

___ [RECORD EXACT NUMBER 0-97]
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF DK OR REF IN Q9A (Q9A=98,99)

Q9B Would it be closer to 0 visits, 1 to 3 visits, 4 to 9 visits, or 10 or more visits?

1 0 visits/None
2 1 to 3
3 4 to 9
4 10 or more
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE WHO VISITED A DOCTOR IN 2010
(EXPERIMENT5=1 and [Q9A=1-97 or Q9B=2-4])

Q10A Which of the following describes how you came up with your answer? Did you think about EACH INDIVIDUAL VISIT; did you think about HOW OFTEN you usually go to the doctor; did you think about TYPES of visits; or did you estimate based on a GENERAL IMPRESSION? [ALLOW MULTIPLE RESPONSES]

1 Think about each visit
2 Think about how often you usually go to the doctor
3 Think about types of visits
4 Estimate based on a general impression
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE WHO VISITED A DOCTOR IN 2010
(EXPERIMENT5=2 and [Q9A=1-97 or Q9B=2-4])

Q10B Which of the following describes how you came up with your answer? Did you estimate based on a GENERAL IMPRESSION; did you think about TYPES of visits; did you think about HOW OFTEN you usually go to the doctor; or did you think about EACH INDIVIDUAL VISIT? [ALLOW MULTIPLE RESPONSES]

1 Estimate based on a general impression
2 Think about types of visits
3 Think about how often you usually go to the doctor
4 Think about each visit
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

END TIMING MODULE 1

START TIMING MODULE 2

SECTION 2: HEALTH INSURANCE

ASK ALL

Q11 The next questions are about health insurance. Include health insurance obtained through employment or purchased directly, as well as government insurance programs like Medicare and Medicaid. Are you covered by any kind of health insurance or health care plan?
[INTERVIEWER NOTE: Health insurance and health care plans include private health insurance, Medicare, Medi-gap, Medicaid, SCHIP/CHIP, military healthcare (TRI-CARE/VA/CHAMP-VA), Indian health service, state-sponsored health plan, other government program, a single service plan (e.g. dental, vision, or prescription)]

1 Yes
2 No
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE WHO ARE INSURED (EXPERIMENT6=1 and Q11=1)

Q12A Would you rate your health insurance as excellent, very good, good, fair, or poor?

1 Excellent
2 Very good
3 Good
4 Fair
5 Poor
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE WHO ARE INSURED (EXPERIMENT6=2 and Q11=1)

Q12B Would you rate your health insurance as poor, fair, good, very good, or excellent?

1 Poor
2 Fair
3 Good
4 Very good
5 Excellent
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF RATED HEALTH INSURANCE (Q12A=1-5 or Q12B=1-5)

Q13 Why do you feel that your health insurance is [INSERT RESPONSE FROM Q12A or Q12B]? [OPEN-END; RECORD VERBATIM RESPONSE]

1 [RECORD VERBATIM RESPONSE]
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q14 The next question is about money that you have spent on medical and dental care for yourself only. Please do NOT count health insurance premiums, over-the-counter drugs, or costs that you were reimbursed for. In 2010, about how much did you spend for medical and dental care? Would you say it was zero dollars... some money but less than $500... $500 to less than $2,000... $2,000 to less than $3,000... $3,000 to less than $5,000... or $5,000 or more?

1 Zero dollars
2 Some money but less than $500
3 $500 to less than $2,000
4 $2,000 to less than $3,000
5 $3,000 to less than $5,000
6 $5,000 or more
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

END TIMING MODULE 2

START TIMING MODULE 3

SECTION 3: POLITICS

[READ TO ALL:] The next questions are about government involvement in health care. Please tell me whether you agree or disagree with the following statements.

ASK RANDOM HALF SAMPLE (EXPERIMENT7=1)

Q15A Increasing government involvement in health care will improve the quality of care. Do you agree or disagree?

1 Agree
2 Disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE (EXPERIMENT7=2)

Q15B Increasing government involvement in health care will hurt the quality of care. Do you agree or disagree?

1 Agree
2 Disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE (EXPERIMENT8=1)

Q16A It is a violation of individual rights for the federal government to require that everyone
have health insurance. Do you agree or disagree?

1 Agree
2 Disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE (EXPERIMENT8=2)

Q16B It is the responsibility of the federal government to require that everyone have health insurance. Do you agree or disagree?

1 Agree
2 Disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

Q17-21 Now I'd like to ask you about some institutions in American society. As I read each one, please tell me how much confidence you have in that institution using a scale from 1 to 10, where 1 means "no confidence at all" and 10 means "great confidence." First, on a scale of 1 to 10, how much confidence do you have in [RANDOMIZE ORDER OF Q17-Q21]? (Next,) how about…[INSERT NEXT ITEM]?

[READ IF NECESSARY: On a scale from 1 to 10, how much confidence do you have in this institution?]

[READ IF NECESSARY: You can use any number between 1 and 10, where 1 means "no confidence at all" and 10 means "great confidence."]

Q17. Congress
Q18. The news media
Q19. The public school system
Q20. The criminal justice system
Q21. The health care system

CATEGORIES

___ [RECORD EXACT NUMBER 1-10]
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE (EXPERIMENT9=1)

Q22A Which of the following do you think is the MOST important thing for Congress to concentrate on right now: the wars in Iraq and Afghanistan, the gap between the rich and the poor, climate change, illegal immigration, or dependence on foreign oil?
[READ IF NECESSARY: If you had to pick from just these 5 choices, which ONE do you think should be the top priority?]

1 The wars in Iraq and Afghanistan
2 The gap between the rich and the poor
3 Climate change
4 Illegal immigration
5 Dependence on foreign oil
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE (EXPERIMENT9=2)

Q22B Which of the following do you think is the MOST important thing for Congress to concentrate on right now: dependence on foreign oil, illegal immigration, climate change, the gap between the rich and the poor, or the wars in Iraq and Afghanistan?
[READ IF NECESSARY: If you had to pick from just these 5 choices, which ONE do you think should be the top priority?]

1 Dependence on foreign oil
2 Illegal immigration
3 Climate change
4 The gap between the rich and the poor
5 The wars in Iraq and Afghanistan
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

[READ TO ALL:] Now I'm going to read a few statements that some people agree with but others disagree with.

ASK RANDOM HALF SAMPLE (EXPERIMENT10=1)

Q23A Do you agree or disagree: Economic growth should be given priority, even if the environment suffers to some extent.

1 Agree
2 Disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE (EXPERIMENT10=2)

Q23B Do you agree or disagree: Protection of the environment should be given priority, even at the risk of slowing economic growth.

1 Agree
2 Disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE (EXPERIMENT11=1)

Q24A Do you agree or disagree: Global warming has been proven.

1 Agree
2 Disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE (EXPERIMENT11=2)

Q24B Do you agree or disagree: Global warming has not been proven.

1 Agree
2 Disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

END TIMING MODULE 3

START TIMING MODULE 4

SECTION 4: EMPLOYMENT

[READ TO ALL:] Now I'm going to ask you some questions about your current employment situation.

ASK ALL

Q25 Last week, were you working full-time, part-time, going to school, keeping house, or what? [CODE ONE RESPONSE ONLY; IF MORE THAN ONE RESPONSE, GIVE PREFERENCE TO FIRST MENTION]

1 Working full-time
2 Working part-time
3 With a job, but not at work because of temporary illness, vacation, strike
4 Unemployed, laid off, looking for work
5 Retired
6 In school
7 Keeping house
8 Other (SPECIFY)
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK EMPLOYED FT/PT OR WITH JOB BUT NOT AT WORK (Q25=1,2,3)

Q26 I have a few questions about your current job. Do you work for a private company, a non-profit organization, or for the government or a government agency?
[INTERVIEWER NOTE: If R has more than 1 job, they should answer about the job where they work the most hours.]

1 Private company
2 A non-profit organization
3 For the government or a government agency
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK EMPLOYED FT/PT OR WITH JOB BUT NOT AT WORK (Q25=1,2,3)

Q27 How many hours a week do you usually work, at all jobs? [INTERVIEWER: If R gives a partial hour (e.g. "15 minutes" or "an hour and a half," please round up to the nearest whole number.]

___ hours [RECORD EXACT NUMBER 0-96]
97 97 hours or more
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

[READ TO ALL:] We are trying to understand how people all over the country are getting along financially, so now I have some questions about earnings and income.

ASK RANDOM HALF SAMPLE (EXPERIMENT12=1)

Q28A In 2010, how much was your total family income, from all sources, before taxes? Total income includes interest or dividends, rent, Social Security, other pensions, alimony or child support, unemployment compensation, public aid or welfare, armed forces or veteran's allotment.
[INTERVIEWER NOTE: If R refuses once, READ: "Information about your income is very important. We greatly appreciate your response and will keep it strictly confidential." IF STILL REFUSED, CODE AS REFUSED. IF R GIVES RANGE, PROBE FOR A DOLLAR AMOUNT.]

_____ dollars [RECORD EXACT NUMBER 0-499,999]
500000 $500,000 or more
777777 (DO NOT READ) Don't know
888888 (DO NOT READ) Refused

ASK IF REFUSED IN Q28A (Q28A=888888)

Q28D1 In 2010, was your total family income from all sources, before taxes, more than $50,000?

1 Yes, more than $50,000
2 No, under $50,000 (incl. exactly $50,000)
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF INCOME IS $50,000 OR LESS (Q28D1=2)

Q28D2 Was it more than $25,000?

1 Yes, more than $25,000
2 No, under $25,000 (incl. exactly $25,000)
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF INCOME IS $25,000 OR LESS (Q28D2=2)

Q28D Was it more than $10,000?

1 Yes, more than $10,000
2 No, under $10,000 (incl. exactly $10,000)
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF INCOME IS MORE THAN $50,000 (Q28D1=1)

Q28D4 Was it more than $75,000?

1 Yes, more than $75,000
2 No, under $75,000 (incl. exactly $75,000)
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM QUARTER SAMPLE (EXPERIMENT12=2)

Q28B1 In 2010, was your total family income from all sources, before taxes, more than $50,000? Total income includes interest or dividends, rent, Social Security, other pensions, alimony or child support, unemployment compensation, public aid or welfare, armed forces or veteran's allotment.
[INTERVIEWER NOTE: If R refuses once, READ: "Information about your income is very important. We greatly appreciate your response and will keep it strictly confidential." IF STILL REFUSED, CODE AS REFUSED]

1 Yes, more than $50,000
2 No, under $50,000 (incl. exactly $50,000)
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF INCOME IS $50,000 OR LESS (Q28B1=2)
Q28B2 Was it more than $25,000?
1 Yes, more than $25,000
2 No, under $25,000 (incl. exactly $25,000)
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF INCOME IS $25,000 OR LESS (Q28B2=2)
Q28B3 Was it more than $10,000?
1 Yes, more than $10,000
2 No, under $10,000 (incl. exactly $10,000)
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF INCOME IS MORE THAN $50,000 (Q28B1=1)
Q28B4 Was it more than $75,000?
1 Yes, more than $75,000
2 No, under $75,000 (incl. exactly $75,000)
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM QUARTER SAMPLE (EXPERIMENT12=3)
Q28C1 In 2010, was your total family income from all sources, before taxes, more than $25,000? Total income includes interest or dividends, rent, Social Security, other pensions, alimony or child support, unemployment compensation, public aid or welfare, armed forces or veteran's allotment.
[INTERVIEWER NOTE: If R refuses once, READ: "Information about your income is very important. We greatly appreciate your response and will keep it strictly confidential." IF STILL REFUSED, CODE AS REFUSED]
1 Yes, more than $25,000
2 No, under $25,000 (incl. exactly $25,000)
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF INCOME IS $25,000 OR LESS (Q28C1=2)
Q28C2 Was it more than $10,000?
1 Yes, more than $10,000
2 No, under $10,000 (incl. exactly $10,000)
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF INCOME IS MORE THAN $25,000 (Q28C1=1)
Q28C3 Was it more than $50,000?
1 Yes, more than $50,000
2 No, under $50,000 (incl. exactly $50,000)
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK IF INCOME IS MORE THAN $50,000 (Q28C3=1)
Q28C4 Was it more than $75,000?
1 Yes, more than $75,000
2 No, under $75,000 (incl. exactly $75,000)
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL
Q29A During 2010, did you receive any income from the following sources: Social Security?
1 Yes
2 No
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL
Q29B How about other retirement or pensions? [READ IF NECESSARY: During 2010, did you receive any income from this source?]
1 Yes
2 No
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL
Q29C How about public assistance or welfare, including Supplemental Security Income, or SSI? [READ IF NECESSARY: During 2010, did you receive any income from this source?]
1 Yes
2 No
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

END TIMING MODULE 4

START TIMING MODULE 5

SECTION 5: CONSENT REQUEST

ASK RANDOM HALF SAMPLE (EXPERIMENT13=1)

CR1A We would like to understand how the use of health care may change as people age. To do that, we need to obtain information about vital statistics, health care costs and diagnoses from your health-related records. In order for us to retrieve these records, we need your consent. This will allow us to conduct more research without asking additional questions. Your consent is voluntary and the information that you provide will be kept completely confidential. May I have your consent to access these records?

1 Consents
2 Declines to consent
98 (DO NOT READ) Don't know

ASK RANDOM HALF SAMPLE (EXPERIMENT13=2)

CR1B We would like to understand how people's income changes as they age. To do that, we need to obtain information about income and employment from your income and employment-related records. In order for us to retrieve these records, we need your consent. This will allow us to conduct more research without asking additional questions. Your consent is voluntary and the information that you provide will be kept completely confidential. May I have your consent to access these records?

1 Consents
2 Declines to consent
98 (DO NOT READ) Don't know

ASK THOSE WHO CONSENTED (CR1A=1 or CR1B=1)

CR2A Can you tell me why you decided to consent to this request to access your records? [OPEN-END; RECORD VERBATIM RESPONSE]

1 [RECORD VERBATIM RESPONSE]
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK THOSE WHO DECLINED TO CONSENT / DK (CR1A=2,98 or CR1B=2,98)

CR2B I appreciate your patience and I indicated that you do NOT consent. Before we move on to the next section, can you tell me why you decided not to consent to this request to access your records? [OPEN-END; RECORD VERBATIM RESPONSE]

1 [RECORD VERBATIM RESPONSE]
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

INT_CR1 INTERVIEWER: DO NOT ASK; Please note any reactions the respondent had to the consent request; Record as many as apply.

1 Hostile
2 Confidentiality concerns
3 Needed clarification
4 Respondent had no reaction
5 Other (SPECIFY)
98 (DO NOT READ) Don't know

END TIMING MODULE 5

START TIMING MODULE 6

SECTION 6: PATRIOTISM

THERE ARE NO Q30 THRU Q35

[READ TO ALL:] Now I am going to read you a few statements. After each one, please tell me how proud you are of America in that area. First...

[RANDOMIZE ORDER OF Q36-Q45]

ASK ALL

Q36 [IF RANDOMIZED 2ND-THRU-LAST, INSERT: How about] The way democracy works.
[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED ITEMS IN THE SERIES, THEN AS NECESSARY: Would you say you are very proud, somewhat proud, not very proud, or not proud at all (of America in this area)?]

1 Very proud
2 Somewhat proud
3 Not very proud
4 Not proud at all
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q37 [IF RANDOMIZED 2ND-THRU-LAST, INSERT: How about] Its political influence in the world.
[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED ITEMS IN THE SERIES, THEN AS NECESSARY: Would you say you are very proud, somewhat proud, not very proud, or not proud at all (of America in this area)?]

1 Very proud
2 Somewhat proud
3 Not very proud
4 Not proud at all
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q38 [IF RANDOMIZED 2ND-THRU-LAST, INSERT: How about] America's economic achievements.
[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED ITEMS IN THE SERIES, THEN AS NECESSARY: Would you say you are very proud, somewhat proud, not very proud, or not proud at all (of America in this area)?]

1 Very proud
2 Somewhat proud
3 Not very proud
4 Not proud at all
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q39 [IF RANDOMIZED 2ND-THRU-LAST, INSERT: How about] Its social security system.
[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED ITEMS IN THE SERIES, THEN AS NECESSARY: Would you say you are very proud, somewhat proud, not very proud, or not proud at all (of America in this area)?]

1 Very proud
2 Somewhat proud
3 Not very proud
4 Not proud at all
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q40 [IF RANDOMIZED 2ND-THRU-LAST, INSERT: How about] Its scientific and technological achievements.
[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED ITEMS IN THE SERIES, THEN AS NECESSARY: Would you say you are very proud, somewhat proud, not very proud, or not proud at all (of America in this area)?]

1 Very proud
2 Somewhat proud
3 Not very proud
4 Not proud at all
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q41 [IF RANDOMIZED 2ND-THRU-LAST, INSERT: How about] Its achievements in sports.
[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED ITEMS IN THE SERIES, THEN AS NECESSARY: Would you say you are very proud, somewhat proud, not very proud, or not proud at all (of America in this area)?]

1 Very proud
2 Somewhat proud
3 Not very proud
4 Not proud at all
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q42 [IF RANDOMIZED 2ND-THRU-LAST, INSERT: How about] Its achievements in the arts and literature.
[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED ITEMS IN THE SERIES, THEN AS NECESSARY: Would you say you are very proud, somewhat proud, not very proud, or not proud at all (of America in this area)?]

1 Very proud
2 Somewhat proud
3 Not very proud
4 Not proud at all
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q43 [IF RANDOMIZED 2ND-THRU-LAST, INSERT: How about] America's armed forces.

[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED ITEMS IN THE SERIES, THEN AS NECESSARY: Would you say you are very proud, somewhat proud, not very proud, or not proud at all (of America in this area)?]

1 Very proud
2 Somewhat proud
3 Not very proud
4 Not proud at all
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q44 [IF RANDOMIZED 2ND-THRU-LAST, INSERT: How about] Its history.

[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED ITEMS IN THE SERIES, THEN AS NECESSARY: Would you say you are very proud, somewhat proud, not very proud, or not proud at all (of America in this area)?]

1 Very proud
2 Somewhat proud
3 Not very proud
4 Not proud at all
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q45 [IF RANDOMIZED 2ND-THRU-LAST, INSERT: How about] Its fair and equal treatment of all groups in society.

[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED ITEMS IN THE SERIES, THEN AS NECESSARY: Would you say you are very proud, somewhat proud, not very proud, or not proud at all (of America in this area)?]

1 Very proud
2 Somewhat proud
3 Not very proud
4 Not proud at all
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

END TIMING MODULE 6

START TIMING MODULE 7

SECTION 7: PRIVACY

[READ TO ALL:] The next questions are about the collection of information by government and businesses.

ASK ALL

Q46 Every ten years, including 2010, most households are sent a Census questionnaire that includes a few questions about everyone living there. Would you agree or disagree that the Census is an invasion of privacy?

1 Agree
2 Disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q47 Do you think the government bothers you too much with requests for information?

1 Yes
2 No
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

[RANDOMIZE ORDER OF Q48-49]

ASK ALL

Q48 [IF RANDOMIZED SECOND, READ: What about your medical records?] How much would it bother you if your medical records were not kept confidential? Would it bother you a lot, some, a little, or not at all?

1 A lot
2 Some
3 A little
4 Not at all
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q49 [IF RANDOMIZED SECOND, READ: What about your income tax records?] How much would it bother you if your income tax records were not kept confidential? Would it bother you a lot, some, a little, or not at all?

1 A lot
2 Some
3 A little
4 Not at all
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q50 Please tell me if you strongly agree, somewhat agree, somewhat disagree, or strongly disagree: People have lost all control over how personal information about them is used.

1 Strongly agree
2 Somewhat agree
3 Somewhat disagree
4 Strongly disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q51 Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?

1 Most people can be trusted
2 You can't be too careful
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q52 Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves?

1 Try to be helpful
2 Just look out for themselves
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q53 Do you think most people would try to take advantage of you if they got a chance, or that they would try to be fair?

1 Would try to take advantage of you
2 Would try to be fair
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q54 How often do you worry about being a victim of identity theft – frequently, occasionally, rarely, or never?

1 Frequently
2 Occasionally
3 Rarely
4 Never
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

END TIMING MODULE 7

START TIMING MODULE 8

SECTION 8: CONSCIENTIOUSNESS

[READ TO ALL:] Now I am going to read a few statements that may or may not describe you. For each statement, please tell me whether you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, or strongly disagree. First…

[RANDOMIZE ORDER OF Q55-Q60D]

ASK ALL

Q55 I am always prepared.
[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED STATEMENTS IN THE SERIES, THEN AS NECESSARY: Do you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, or strongly disagree?]

1 Strongly agree
2 Somewhat agree
3 Neither agree nor disagree
4 Somewhat disagree
5 Strongly disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q56 I carry out my plans.
[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED STATEMENTS IN THE SERIES, THEN AS NECESSARY: Do you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, or strongly disagree?]

1 Strongly agree
2 Somewhat agree
3 Neither agree nor disagree
4 Somewhat disagree
5 Strongly disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q57 I pay attention to details.
[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED STATEMENTS IN THE SERIES, THEN AS NECESSARY: Do you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, or strongly disagree?]

1 Strongly agree
2 Somewhat agree
3 Neither agree nor disagree
4 Somewhat disagree
5 Strongly disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q58 I waste my time.
[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED STATEMENTS IN THE SERIES, THEN AS NECESSARY: Do you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, or strongly disagree?]

1 Strongly agree
2 Somewhat agree
3 Neither agree nor disagree
4 Somewhat disagree
5 Strongly disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q59 I do just enough work to get by.
[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED STATEMENTS IN THE SERIES, THEN AS NECESSARY: Do you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, or strongly disagree?]

1 Strongly agree
2 Somewhat agree
3 Neither agree nor disagree
4 Somewhat disagree
5 Strongly disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q60 I don't see things through.
[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED STATEMENTS IN THE SERIES, THEN AS NECESSARY: Do you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, or strongly disagree?]

1 Strongly agree
2 Somewhat agree
3 Neither agree nor disagree
4 Somewhat disagree
5 Strongly disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q60A I make plans and stick to them.
[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED STATEMENTS IN THE SERIES, THEN AS NECESSARY: Do you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, or strongly disagree?]

1 Strongly agree
2 Somewhat agree
3 Neither agree nor disagree
4 Somewhat disagree
5 Strongly disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q60B I have difficulty getting started doing work.
[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED STATEMENTS IN THE SERIES, THEN AS NECESSARY: Do you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, or strongly disagree?]

1 Strongly agree
2 Somewhat agree
3 Neither agree nor disagree
4 Somewhat disagree
5 Strongly disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q60C I avoid my duties.
[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED STATEMENTS IN THE SERIES, THEN AS NECESSARY: Do you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, or strongly disagree?]

1 Strongly agree
2 Somewhat agree
3 Neither agree nor disagree
4 Somewhat disagree
5 Strongly disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q60D I get chores done right away.
[INTERVIEWER: READ FOR FIRST 2 RANDOMIZED STATEMENTS IN THE SERIES, THEN AS NECESSARY: Do you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, or strongly disagree?]

1 Strongly agree
2 Somewhat agree
3 Neither agree nor disagree
4 Somewhat disagree
5 Strongly disagree
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

END TIMING MODULE 8

START TIMING MODULE 9

SECTION 9: BACKGROUND

[READ TO ALL:] And now just a few background questions.

ASK ALL

QTV LAST WEEK, how many hours did you spend watching television?
[INTERVIEWER: If R gives a partial hour (e.g. "15 minutes" or "an hour and a half"), please round up to the nearest whole number.]

____ [RECORD EXACT NUMBER 0-168]
998 (DO NOT READ) Don't know
999 (DO NOT READ) Refused

ASK IF QTV=1-168

INT_TV INTERVIEWER: DO NOT ASK; CODE RESPONDENT'S APPROACH TO ANSWERING QUESTION: IF R DID NOT SAY HOW, RECORD AS "DON'T KNOW"; RECORD AS MANY AS APPLY

1 Thought about specific days of the week and added them up (e.g., 2 hours Monday plus 3 hours Friday)
2 Thought about how many hours usually watch per day and used that as a point of reference (e.g., I usually watch 2 hours a day)
3 Thought about types of shows (e.g., news, movies) and added them up
4 Other (SPECIFY)
98 (DO NOT READ) Don't know

[RANDOMIZE Q61A1/Q61B1 WITH Q62A1/Q62B1; RANDOM HALF WILL GET Q61A1/Q61B1 FIRST (EXPERIMENT17=1) AND RANDOM HALF WILL GET Q62A1/Q62B1 FIRST (EXPERIMENT17=2)]

ASK RANDOM HALF SAMPLE (EXPERIMENT14=1)

Q61A1 In a TYPICAL week, how many hours do you spend using a computer?
[INTERVIEWER NOTE: Accept responses in hours or in minutes; in Q61A2, note whether response was reported in hours or minutes.]

____ [RECORD EXACT NUMBER 0-9997]
9998 (DO NOT READ) Don't know
9999 (DO NOT READ) Refused

ASK IF SPECIFIED COMPUTER TIME (Q61A1=0-9997)

Q61A2 INTERVIEWER: If R already stated that time spent on computer was in hours or minutes, do not ask and enter 1 or 2. Otherwise, ASK: Would you say that time is in hours or minutes?
[PROGRAMMER: Auto-punch Q61A2=3 (not applicable) when Q61A1=0]

1 Hours
2 Minutes
3 (DO NOT READ) Not applicable [PROGRAMMER: Punch 3 only for Q61A1=0]
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE (EXPERIMENT14=2)

Q61B1 In a TYPICAL week, how many hours do you spend using a computer? Please do not include any time spent writing or reading emails.
[INTERVIEWER NOTE: Accept responses in hours or in minutes; in Q61B2, note whether response was reported in hours or minutes.]

____ [RECORD EXACT NUMBER 0-9997]
9998 (DO NOT READ) Don't know
9999 (DO NOT READ) Refused

ASK IF SPECIFIED COMPUTER TIME (Q61B1=0-9997)

Q61B2 INTERVIEWER: If R already stated that time spent on computer was in hours or minutes, do not ask and enter 1 or 2. Otherwise, ASK: Would you say that time is in hours or minutes?
[PROGRAMMER: Auto-punch Q61A2=3 (not applicable) when Q61B1=0]

1 Hours
2 Minutes
3 (DO NOT READ) Not applicable [PROGRAMMER: Punch 3 only for Q61B1=0]
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE (EXPERIMENT15=1)

Q62A1 In a TYPICAL week, how many hours do you spend talking on the telephone?
[INTERVIEWER NOTE: Accept responses in hours or in minutes; in Q62A2, note whether response was reported in hours or minutes.]

____ [RECORD EXACT NUMBER 0-9997]
9998 (DO NOT READ) Don't know
9999 (DO NOT READ) Refused

ASK IF SPECIFIED TELEPHONE TIME (Q62A1=0-9997)

Q62A2 INTERVIEWER: If R already stated that time spent talking on the telephone was in
hours or minutes, do not ask and enter 1 or 2. Otherwise, ASK: Would you say that time is in hours or minutes?
[PROGRAMMER: Auto-punch Q62A2=3 (not applicable) when Q62A1=0]

1 Hours
2 Minutes
3 (DO NOT READ) Not applicable [PROGRAMMER: Punch 3 only for Q62A1=0]
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK RANDOM HALF SAMPLE (EXPERIMENT15=2)

Q62B1 In a TYPICAL week, how many hours do you spend talking on the telephone? Please do
not include time spent speaking with family members.
[INTERVIEWER NOTE: Accept responses in hours or in minutes; in Q62B2, note whether response was reported in hours or minutes.]

____ [RECORD EXACT NUMBER 0-9997]
9998 (DO NOT READ) Don't know
9999 (DO NOT READ) Refused

ASK IF SPECIFIED TELEPHONE TIME (Q62B1=0-9997)

Q62B2 INTERVIEWER: If R already stated that time spent talking on the telephone was in hours or minutes, do not ask and enter 1 or 2. Otherwise, ASK: Would you say that time is in hours or minutes?
[PROGRAMMER: Auto-punch Q62B2=3 (not applicable) when Q62B1=0]

1 Hours
2 Minutes
3 (DO NOT READ) Not applicable [PROGRAMMER: Punch 3 only for Q62B1=0]
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

THERE IS NO Q63

ASK ALL

Q64 How many years have you been living in your current home? [INTERVIEWER: IF R says "All my life", PROBE FOR NUMBER OF YEARS]

___ years [RECORD EXACT NUMBER 0-97]
0 Less than 1 year
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

THERE IS NO Q65

ASK ALL

Q66 In what month and year were you born?

1 [RECORD RESPONSE IN THIS FORMAT: MM/YYYY]
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q67 Are you Spanish, Hispanic, or Latino?

1 Yes
2 No
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q68 I am going to read you a list of five race categories. Please choose one or more races that you consider yourself to be: White; Black or African-American; American Indian or Alaska Native; Asian; OR Native Hawaiian or Other Pacific Islander.
[INTERVIEWER NOTE: DO NOT PROBE UNLESS RESPONSE IS HISPANIC OR A HISPANIC ORIGIN; ALLOW MULTIPLE RESPONSES IF R SAYS HISPANIC OR LATINO, PROBE: Do you consider yourself a WHITE (Hispanic/Latino) or a BLACK (Hispanic/Latino)? IF R DOES NOT SAY WHITE, BLACK OR ONE OF THE RACE CATEGORIES LISTED, RECORD AS "OTHER" (CODE 6)]

1 White
2 Black or African-American
3 American Indian or Alaska native
4 Asian
5 Native Hawaiian or other pacific islander
6 Other
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q69 What is the highest level of school you have completed or the highest degree you have received? [DO NOT READ BUT CAN PROBE FOR CLARITY IF NECESSARY]

1 Less than high school
2 High school graduate, High school diploma or the equivalent (for example: GED)
3 Some college but no degree
4 Associate degree
5 Bachelor's degree (for example: B.A., A.B., B.S.)
6 Graduate degree [master's degree, professional school degree, or doctorate degree]
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

ASK ALL

Q70 What state do you currently live in?

1	Alabama	28	Nebraska
2	Alaska	29	Nevada
3	Arizona	30	New Hampshire
4	Arkansas	31	New Jersey
5	California	32	New Mexico
6	Colorado	33	New York
7	Connecticut	34	North Carolina
8	Delaware	35	North Dakota
9	District of Columbia	36	Ohio
10	Florida	37	Oklahoma
11	Georgia	38	Oregon
12	Hawaii	39	Pennsylvania
13	Idaho	40	Rhode Island
14	Illinois	41	South Carolina
15	Indiana	42	South Dakota
16	Iowa	43	Tennessee
17	Kansas	44	Texas
18	Kentucky	45	Utah
19	Louisiana	46	Vermont
20	Maine	47	Virginia
21	Maryland	48	Washington State
22	Massachusetts	49	West Virginia
23	Michigan	50	Wisconsin
24	Minnesota	51	Wyoming
25	Mississippi	98	(DO NOT READ) Don’t know
26	Missouri	99	(DO NOT READ) Refused
27	Montana		

ASK ALL (DO NOT ASK IF PRETEST)

Q71 A letter describing this study may have been sent to your home recently. Do you remember seeing the letter?

1 Yes
2 No
98 (DO NOT READ) Don’t know
99 (DO NOT READ) Refused

ASK IF SAW LETTER (Q71=1)

Q72 Do you happen to remember if there was anything else in the envelope with the letter? [IF YES AND SAID ANY AMOUNT OF MONEY WITHOUT PROMPTING, ENTER CODE=1; IF YES AND DID NOT SPECIFY, PROBE: Could you please tell me what was included with the letter?]

1 Yes, money
2 Yes, something other than money (SPECIFY)
3 No, nothing was included with the letter
98 (DO NOT READ) Don't know
99 (DO NOT READ) Refused

END TIMING MODULE 9

[THANK AND END INTERVIEW:] These are all of the questions we have for you. Thank you very much for your time. Good-bye.

POST-INTERVIEW OBSERVATIONS TO BE ANSWERED BY INTERVIEWER

Q73 INTERVIEWER, PLEASE ANSWER: The respondent answered the survey questions to the best of his or her ability.

1 Not at all
2 Not that often
3 Somewhat often
4 Pretty often
5 Very often

Q74 INTERVIEWER, PLEASE ANSWER: The respondent was reluctant to answer the survey questions.

1 Not at all
2 Not that often
3 Somewhat often
4 Pretty often
5 Very often

Q75 INTERVIEWER, PLEASE ANSWER: The respondent had trouble understanding the survey questions.

1 Not at all
2 Not that often
3 Somewhat often
4 Pretty often
5 Very often

APPENDIX B

SURVEY ITEMS USED FOR EACH EFFORT INDICATOR (CHAPTER 2)

Indicator	Items
Item nonresponse	All items
Responses to open-ended items	Q2, Q13, CR2A/B
Straight-lining / non-differentiation	Q17-Q21, Q36-Q46, Q55-Q60D
Acquiescence	Q6A1, Q6B1, Q6C1, Q6D1, Q6E1, Q6F1, Q7, Q11, Q15A/B, Q16A/B, Q23A/B, Q24A/B, Q29A-C, CR1A/B, Q46, Q47, Q55-Q06d
Response order effects	Q1A/B, Q3A/B, Q12A/B, Q14, Q22, Q36-Q45, Q50, Q54-Q60D
Providing a round value for a numerical response	Q4A1/B1, Q5A1/B1, Q6A2, Q6B2, Q6C2, Q6D2, Q6E2, Q6F2, Q9A, Q27, Q28A, QTV, Q61A1/B1M Q62A1/B1, Q64
Estimation strategy for a numerical response	Respondent: Q10A/B; Interviewer: INT_4, INT_5, INT_TV
Lack of attention to question wording	Q4A1/B1, Q5A1/B1, Q61A1/B1, Q62A1/B1
Underreporting to filter items	Q6A1, Q6B1, Q6C1, Q6D1 Q6E1, Q6F1
Accuracy of reports as compared to frame information	Q28, Q64
Interviewer report of respondent effort	Q73
Age	Q66
Education	Q69
Interviewer report of respondent difficulty	Q75
Conscientiousness	Q55-Q60D
Incentive recall	Q71, Q72

APPENDIX C

MARYLAND MAIL SURVEY MATERIALS

C1. ADVANCE LETTER

[DATE]
[ADDRESS]

Dear [NAME],

I am writing to ask for your help with an important study being conducted by researchers at the University of Maryland. The results of this study will help us learn more about the health and daily life of people living in Maryland. You have been selected for this study as part of a scientific sample.

A few days from now you will receive a questionnaire in the mail for the 2011 Maryland Survey on Health and Lifestyles. This survey will ask questions about your daily life and collect your opinions on important issues. [INCENTIVE: You will also receive a small token of our appreciation for your help.]

When the questionnaire arrives, we hope you will take a few minutes to help us by filling it out and mailing it back promptly. The success of this research depends on the generous help of people like you.

Thank you in advance for your help.

Sincerely,

Dr. Roger Tourangeau
Project Director
University of Maryland

C2. INITIAL COVER LETTER

[DATE]
[ADDRESS]

Dear [NAME],

We recently sent you a letter about an important study being conducted by researchers at the University of Maryland. The results of this study will help us learn more about the health and daily life of people living in Maryland. You have been selected for this study as part of a scientific sample.

We have enclosed a questionnaire for the 2011 Maryland Survey on Health and Lifestyles. This survey asks questions about your daily life and collects your opinions on important issues. [INCENTIVE: We have also included a small token of appreciation for your participation.]

Your participation is voluntary, and you may skip any questions you prefer not to answer. All of your responses will be kept confidential and will be used for research purposes only.

If you have any questions or comments about this survey, we would be happy to hear from you. You can email your questions to info@MarylandSurvey.net.

Thank you in advance for your participation. We look forward to receiving your responses.

Sincerely,

Dr. Roger Tourangeau
Project Director
University of Maryland

C3. THANK YOU / REMINDER POSTCARD

We recently sent you a questionnaire for the 2011 Maryland Survey on Health and Lifestyles. The results of this study will provide valuable information to researchers at the University of Maryland about the health and daily life of people living in Maryland.

We have not received your response yet. If you have already responded, thank you. If you have not responded yet, please complete the questionnaire and mail it back in the postage-paid envelope as soon as possible. If you did not receive the questionnaire or if yours was misplaced, please contact us at info@MarylandSurvey.net, and we will send you a new copy.

Thank you,

Dr. Roger Tourangeau
Project Director

C4. REPLACEMENT COVER LETTER

[DATE]
[ADDRESS]

Dear [NAME],

We recently sent you a questionnaire for the 2011 Maryland Survey on Health and Lifestyles. The results of this study will help researchers at the University of Maryland learn more about the health and daily life of people living in Maryland.

We have not received your response yet. If you have already responded, thank you. If you have not responded yet, we have included another questionnaire with this letter. Please complete the questionnaire and mail it back in the postage-paid envelope as soon as possible.

As mentioned before, your participation is voluntary, and you may skip any questions you prefer not to answer. All of your responses will be kept confidential and will be used for research purposes only.

If you have any questions or comments about this survey, we would be happy to hear from you. You can email your questions to info@MarylandSurvey.net.

Thank you in advance for your participation. We look forward to receiving your responses.

Sincerely,

Dr. Roger Tourangeau
Project Director
University of Maryland

C5. QUESTIONNAIRE

WE MAILED THIS PACKAGE TO A SPECIFIC PERSON AT THIS ADDRESS. IT IS IMPORTANT THAT ONLY THIS PERSON ANSWER THE QUESTIONS.

Q1 Would you say your health in general is excellent, very good, good, fair, or poor?
1 Excellent
2 Very good
3 Good
4 Fair
5 Poor

Q2 During the past 12 month, how many times did you see a doctor or other health care professional about your health at a doctor's office, clinic, hospital emergency room, at home, or some other place?
1 None
2 Once
3 2 to 5 times
4 6 to 9 times
5 10 or more times

Q3 Has a doctor or other health care professional ever told you that you had any of the following?
1 Yes
2 No

Q3A Asthma
Q3B Diabetes
Q3C Arthritis
Q3D Hypertension or high blood pressure
Q3E A sexually transmitted disease (STD)
Q3F Psoriasis

Q4 During the past 12 months, how many times did you visit a dentist?
1 None
2 Once
3 Twice
4 3 times or more

Q5 Do you consider yourself to be overweight, underweight, or about the right weight?
1 Overweight
2 Underweight
3 About the right weight

Q6 In the past 7 days, how often did you eat breakfast?

1 No days
2 One day
3 A few days
4 Most days
5 Every day

Q7 In the past 7 days, how often did you eat or drink each of the following?
1 No days
2 One day
3 A few days
4 Most days
5 Every day

Q7A Milk
Q7B Soda or pop
Q7C Vegetables
Q7D Pasta
Q7E Chicken
Q7F Cookies or cake

Q8 Did you take vitamins, minerals, herbal medicine, or other dietary supplements in the past 7 days?
1 Yes
2 No

Q9 Did you take nonprescription painkillers, such as ibuprofen, or aspirin, in the past 7 days?
1 Yes
2 No

Q10 Did you use dental floss in the past 7 days?
1 Yes
2 No

Q11 Did you participate in any moderate-intensity activities that cause a small increase in breathing or heart rate, such as brisk walking, swimming, or bicycling, in the past 7 days?
1 Yes
2 No

Q12 Did you participate in any vigorous-intensity activities that cause a large increase in breathing or heart rate, such as running or basketball, in the past 7 days?
1 Yes
2 No

Q13 Over the past 2 weeks, how often have you felt tired or had little energy?

1 No days
2 A few days
3 More than half the days
4 Nearly every day

Q14 Over the past 2 weeks, how often have you had little interest or pleasure in doing things?
1 No days
2 A few days
3 More than half the days
4 Nearly every day

Q15 Over the past 2 weeks, how often have you had trouble concentrating on things?
1 No days
2 A few days
3 More than half the days
4 Nearly every day

Q16 Over the past 2 weeks, how often have you felt down, depressed, or hopeless?
1 No days
2 A few days
3 More than half the days
4 Nearly every day

Q17 In the past 30 days, on how many days did you drink one or more drinks of an alcoholic beverage? *Count as a drink a can or bottle of beer, a wine cooler or a glass of wine, a glass of champagne or sherry, a shot of liquor, or a mixed drink or cocktail.*
____ drinks

IF NONE, GO TO ITEM 19.

Q18 In the past 30 days, were there any days when you had 5 or more drinks on the same occasion?
1 Yes
2 No

Q19 Have you ever, even once, smoked a cigarette?
1 Yes
2 No

Q20 In the past 30 days, did you smoke cigarettes every day, some days, or not at all?
1 Every day
2 Some days
3 Not at all

Q21 Have you ever, even once, used marijuana or hashish?

1 Yes
2 No

Q22 Have you used marijuana or hashish in the past 12 months?
1 Yes
2 No

THIS NEXT SET OF QUESTIONS ASKS ABOUT THINGS YOU MAY DO IN YOUR FREE TIME.

Q23 How many days per week do you typically read the newspaper?
1 No days
2 One day
3 A few days
4 Most days
5 Every day

Q24 Did you read any books in the past 30 days?
1 Yes
2 No

Q25 Over the past 30 days, how many hours per day did you typically watch television or movies?
___ hours

Q26 Were you able to devote any time to volunteer work in the past 12 months?
1 Yes
2 No

THIS NEXT SET OF QUESTIONS ASKS ABOUT YOUR LEVEL OF INTEREST AND INVOLVEMENT IN POLITICS.

Q27 How interested are you in what's going on in politics and the government?
1 Not interested at all
2 Slightly interested
3 Moderately interested
4 Very interested
5 Extremely interested

Q28 To what extent do you agree or disagree with the following statements?
1 Strongly disagree
2 Somewhat disagree
3 Somewhat agree
4 Strongly agree

Q28A Increasing government involvement in health care would improve the quality of care.
Q28B It would be a violation of individual rights for the federal government to require that everyone has health insurance.

Q29 Please think back to the last GENERAL election, which was the midterm election in November 2010. Do you remember whether or not you voted in that election?
1 Yes, I voted in 2010
2 No, I did not vote in 2010

Q30 In the 2008 PRESIDENTIAL election, John McCain ran on the Republican ticket against Barack Obama on the Democratic ticket. Do you remember whether or not you voted in that election?
1 Yes, I voted in 2008
2 No, I did not vote in 2008

Q31 In the 2004 PRESIDENTIAL election, George W. Bush ran on the Republican ticket against John Kerry on the Democratic ticket. Do you remember whether or not you voted in that election?
1 Yes, I voted in 2004
2 No, I did not vote in 2004

Q32 To what extent do you agree or disagree with the following statements?
1 Strongly disagree
2 Somewhat disagree
3 Somewhat agree
4 Strongly agree

Q32A Men are better suited emotionally for politics than women.
Q32B A working mother can establish just as warm and secure a relationship with her children as a mother who does not work.

THIS NEXT SET OF STATEMENTS MAY OR MAY NOT DESCRIBE YOU. PLEASE INDICATE HOW TRUE EACH STATEMENT IS.

Q33 I sometimes drive faster than the speed limit.
1 Very true
2 Somewhat true
3 Not very true
4 Not true at all

Q34 I never cover up my mistakes.
1 Very true
2 Somewhat true
3 Not very true
4 Not true at all

Q35 I sometimes tell lies if I have to.
1 Very true
2 Somewhat true
3 Not very true
4 Not true at all

Q36 I never swear.
1 Very true
2 Somewhat true
3 Not very true
4 Not true at all

Q37 I have never dropped litter on the street.
1 Very true
2 Somewhat true
3 Not very true
4 Not true at all

Q38 There have been occasions when I have taken advantage of someone.
1 Very true
2 Somewhat true
3 Not very true
4 Not true at all

Q39 I have taken sick leave from work or school even though I wasn't really sick.
1 Very true
2 Somewhat true
3 Not very true
4 Not true at all

Q40 I never take things that don't belong to me.
1 Very true
2 Somewhat true
3 Not very true
4 Not true at all

Q41 I have said something bad about a friend behind his or her back.
1 Very true
2 Somewhat true
3 Not very true
4 Not true at all

Q42 When I hear people talking privately, I avoid listening.
1 Very true
2 Somewhat true
3 Not very true
4 Not true at all

LASTLY, JUST A FEW BACKGROUN QUESTIONS.

Q43 What is your gender?
1 Female
2 Male

Q44 In what month and year were you born?

Q45 What is the highest level of education you have completed?
1 Less than high school
2 High school graduate
3 Some college but no degree
4 Associate's degree
5 Bachelor's degree
6 Graduate degree

Q46 Are you Spanish, Hispanic, or Latino?
1 Yes
2 No

Q47 Which of the following races do you consider yourself to be?
Mark all that apply.
1 White
2 Black or African American
3 American Indian or Alaska Native
4 Asian
5 Native Hawaiian or other Pacific Islander

Q48 What is your marital status?

1 Married
2 Divorced
3 Separated
4 Widowed
5 Never married

Q49 In 2010, what was your total household income from all sources, before taxes?
1 Less than $25,000
2 $25,000 to $49,999
3 $50,000 to $74,999
4 $75,000 to $99,999
5 $100,000 or more

Q50 Aside from weddings and funerals, how often do you attend religious services?
1 Once a week or more
2 Once or twice a month
3 A few times a year
4 Once a year
5 Never

ONE GOAL OF THIS SURVEY IS TO FIND WAYS TO IMPROVE FUTURE SURVEYS. WE WOULD LIKE TOHEAR YOUR OPINIONS ABOUT TOPICS ASKED ABOUT IN THIS SURVEY.

Q51 To what extent do you think that people you know might falsely report their answers to questions on the following topics?
1 Not at all
2 Some
3 A lot

Q51A Medical conditions
Q51B Diet
Q51C Exercise
Q51D Mental health
Q51E Alcohol use
Q51F Smoking
Q51G Drug use
Q51H Volunteering
Q51I Voting
Q51J Income
Q51K Religious service attendance

THANK YOU. WE APPRECIATE YOUR PARTICIPATION IN THIS STUDY!

APPENDIX D

ITEM SENSITIVITY RATINGS (CHAPTER 3)

I collected three indicators of item sensitivity: respondent ratings, item nonresponse rates, and coder ratings. In this Appendix, I provide information about each of these indicators.

Respondent ratings. Respondents utilized a three-point scale to rate how likely they thought people would be to falsely report their answers to questions on 11 topics. These ratings are provided in Table D.1. Respondents rated drug use and alcohol use as the most sensitive topics, and they rated volunteering and voting as the least sensitive topics. The rank-order of the topics was identical for the incentive and control groups. I estimated a linear model for each item in which the independent variables were incentive receipt and a 6-category age variable. Controlling for age, incentive group respondents rated 10 of the 11 topics as being less sensitive than the control group, although only three of these differences were significant at the 0.05 level (alcohol use (2.00 and 2.12, respectively), medical conditions (1.68 and 1.76, respectively), and volunteering (1.46 and 1.55, respectively)).

Item nonresponse. Item nonresponse ranged from to 0.1% (health status) to 25.5% (number of days drank alcohol in past 30 days).[22] The median proportion of respondents skipping a particular item was 1.7%, while the mean proportion was 4.3%. Item nonresponse appeared to be driven by burden more so than sensitivity. Eight of the ten items with the highest missing rates were either open-ended items or part of grids.

[22] In order to be comparable to respondent ratings, the analysis was restricted to respondents who answered the age question.

The two exceptions were: income (8.9%) and having used marijuana in the past 12 months (6.3%).

Table D.1. Mean Respondent Sensitivity Ratings

	Overall Mean (95% CI)
Drug use	2.06 (2.01, 2.12)
Alcohol use	2.02 (1.98, 2.07)
Diet	1.97 (1.92, 2.01)
Exercise	1.92 (1.88, 1.97)
Mental health	1.90 (1.85, 1.94)
Income	1.80 (1.75, 1.84)
Smoking	1.77 (1.72, 1.81)
Medical conditions	1.71 (1.68, 1.75)
Religious service attendance	1.58 (1.54, 1.62)
Volunteering	1.50 (1.47, 1.55)
Voting	1.49 (1.44, 1.52)
Sample size[1]	932 – 953

1. Range in sample size due to item nonresponse. Analysis restricted to respondents who answered age question.

Coder ratings. Twenty coders rated the sensitivity of each of 45 survey items on a five-point scale. The mean ratings ranged from 1.10 to 4.60. The five most sensitive items were having been diagnosed with an STD (4.60), having used marijuana in the past 12 months (4.25), having had five or more alcoholic drinks in a given day in the past month (3.75), ever having used marijuana (3.60), and being overweight (3.35). The five least sensitive items were frequency of eating chicken in the past seven days (1.10), frequency of eating pasta in the past seven days (1.20), taking nonprescription painkillers in the past seven days (1.40), taking vitamins in the past seven days (1.45), and frequency of drinking milk in the past seven days (1.50).

Choosing an indicator of item sensitivity. Each method had pros and cons. The benefit of respondent ratings and item nonresponse was that these pieces of information were provided by the respondents themselves. However, the respondent ratings were based on broad topics, as opposed to individual items. Additionally, item nonresponse also can be an indicator of respondent burden or fatigue; in fact, eight of the ten items with the highest nonresponse rates were the types of items that are typically considered more burdensome, such as open-ended items or those that are presented in grids. Additionally, respondents who feel an item is sensitive may still feel obligated answer it and to deal with their discomfort by providing a falsified response.

Considering these pros and cons, I decided to use the coder ratings as the main indicator of item sensitivity. These ratings were available for all 45 items. They also seemed to correlate rather well with the other sensitivity ratings. For example, respondents rated drug use and alcohol use as the two most sensitive topics, and the coders rated three of the four items on these topics as being among the most sensitive

items in the survey. Two of the four items on these topics were also among the most commonly skipped items. There was also a positive correlation between the 45 coder sensitivity ratings and item nonresponse (r=0.172), although it was not statistically different from zero (p=0.259).

As discussed above, I used coder ratings as the main indicator of item sensitivity. Based on the mean coder ratings, I placed the items into three equally-sized groups that I referred to as high, medium, and low sensitivity (Table D.2).

Table D.2. Grouping Items by Sensitivity

Sensitivity	**Survey Items**
High	Q1, Q3b, Q3e, Q5, Q16, Q17, Q18, Q20, Q21, Q22, Q30, Q31, Q32a, Q32b, Q49
Medium	Q2, Q4, Q3d, Q3f, Q7c, Q7f, Q10, Q11, Q12, Q14, Q15, Q19, Q26, Q29, Q50
Low	Q3a, Q3c, Q6, Q7a, Q7b, Q7d, Q7e, Q8, Q9, Q13, Q23, Q24, Q27, Q28a, Q28b

APPENDIX E

CONSUMER SENTIMENT ITEMS

PAGO We are interested in how people are getting along financially these days. Would you say that you (and your family living there) are better off or worse off financially than you were a year ago?

1 BETTER NOW
3 SAME
5 WORSE NOW
8 DK
9 NA

PEXP Now looking ahead--do you think that a year from now you (and your family living there) will be better off financially, or worse off, or just about the same as now?

1 WILL BE BETTER OFF
3 SAME
5 WILL BE WORSE OFF
8 DK
9 NA

BUS12 Now turning to business conditions in the country as a whole--do you think that during the next 12 months we'll have good times financially, or bad times, or what?

1 GOOD TIMES
2 GOOD WITH QUALIFICATIONS
3 PRO-CON
4 BAD WITH QUALIFICATIONS
5 BAD TIMES
8 DK
9 NA

BUS5 Looking ahead, which would you say is more likely -- that in the country as a whole we'll have continuous good times during the next 5years or so, or that we will have periods of widespread unemployment or depression, or what?

1 (Continuous) good times; boom; prosperity; no recession
2 Good times, qualified (not bad); pretty good, no unemployment, no depression
3 Pro-con; some recession, some unemployment, periods of unemployment
4 Bad times, qualified (not good); recession; bad at some times but not most of the time; periods of widespread unemployment; some depression; unemployment
5 Bad times, depression; widespread unemployment
98 DK; can't tell
99 NA; R speaks only of hopes and wishes; R gives only comparative or relative answer, "Better," "Same," "Worse"; "more/less unemployment or inflation"

DUR About the big things people buy for their homes -- such as furniture, a refrigerator, stove, television, and things like that. Generally speaking, do you think now is a good or a bad time for people to buy major household items?

1 GOOD
3 PRO-CON
5 BAD
8 DK
9 NA

APPENDIX F

ADDITIONAL ANALYSES FOR CHAPTER 4

F1. CORRELATION MATRICES

Correlations above diagonal are for incentive condition respondents; correlations below diagonal are for control condition respondents.

Table F.1A. Correlation Matrix for Patriotism Items

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 1	--	0.20	0.31	0.13	0.22	0.18	0.14	0.22	0.28	0.32
Item 2	0.42	--	0.33	0.13	0.27	0.19	0.19	0.17	0.20	0.21
Item 3	0.46	0.34	--	0.20	0.30	0.26	0.14	0.12	0.20	0.17
Item 4	0.26	0.22	0.21	--	0.14	0.17	0.18	0.03	0.05	0.12
Item 5	0.35	0.28	0.33	0.14	--	0.29	0.21	0.13	0.20	0.12
Item 6	0.31	0.27	0.30	0.24	0.32	--	0.23	0.21	0.21	0.12
Item 7	0.38	0.31	0.36	0.24	0.40	0.37	--	0.02	0.18	0.00
Item 8	0.23	0.14	0.18	0.14	0.31	0.36	0.26	--	0.36	0.25
Item 9	0.41	0.30	0.37	0.15	0.40	0.31	0.37	0.35	--	0.30
Item 10	0.42	0.31	0.30	0.11	0.27	0.33	0.27	0.24	0.39	--

Table F.1B. Correlation Matrix for Conscientiousness Items

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 1	--	0.34	0.27	0.23	0.18	0.16	0.37	0.25	0.10	0.33
Item 2	0.31	--	0.23	0.25	0.17	0.23	0.48	0.29	0.11	0.32
Item 3	0.30	0.35	--	0.13	0.22	0.16	0.32	0.14	0.12	0.23
Item 4	0.24	0.22	0.20	--	0.29	0.28	0.27	0.29	0.24	0.24
Item 5	0.14	0.17	0.10	0.42	--	0.39	0.23	0.45	0.36	0.25
Item 6	0.11	0.23	0.20	0.27	0.29	--	0.20	0.36	0.35	0.13
Item 7	0.36	0.52	0.35	0.15	0.14	0.20	--	0.26	0.15	0.32
Item 8	0.22	0.14	0.12	0.38	0.32	0.29	0.18	--	0.32	0.32
Item 9	0.23	0.13	0.14	0.29	0.24	0.27	0.14	0.29	--	0.15
Item 10	0.35	0.37	0.30	0.29	0.21	0.23	0.40	0.33	0.27	--

Table F.1C. Correlation Matrix for Impression Management Items

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 1	--	-0.11	0.22	0.22	0.05	0.16	0.19	-0.07	0.20	0.10
Item 2	0.00	--	0.09	0.03	0.13	0.09	-0.01	0.19	0.02	0.10
Item 3	0.28	0.16	--	0.17	0.15	0.38	0.34	0.19	0.26	0.24
Item 4	0.21	0.19	0.17	--	0.19	0.17	0.24	0.06	0.25	0.24
Item 5	0.08	0.22	0.05	0.22	--	0.22	0.22	0.27	0.24	0.19
Item 6	0.24	0.18	0.39	0.15	0.22	--	0.36	0.14	0.45	0.22
Item 7	0.27	0.04	0.33	0.21	0.13	0.27	--	0.09	0.26	0.19
Item 8	-0.12	0.16	0.02	0.04	0.19	0.12	-0.20	--	0.11	0.18
Item 9	0.25	0.08	0.34	0.25	0.14	0.29	0.34	0.01	--	0.21
Item 10	0.15	0.12	0.21	0.19	0.14	0.17	0.14	0.15	0.22	--

Table F.1D. Correlation Matrix for Consumer Sentiment Items: $10 vs. Control

	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	--	0.11	0.25	0.27	0.16
Item 2	0.21	--	0.42	0.39	0.12
Item 3	0.43	0.26	--	0.65	0.20
Item 4	0.43	0.28	0.52	--	0.19
Item 5	0.11	-0.10	0.11	0.13	--

Table F.1E. Correlation Matrix for Consumer Sentiment Items: $5 vs. Control

	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	--	0.10	0.15	0.17	0.13
Item 2	0.21	--	0.33	0.25	0.09
Item 3	0.43	0.26	--	0.60	0.27
Item 4	0.43	0.28	0.52	--	0.15
Item 5	0.11	-0.10	0.11	0.13	--

Table F.1F. Correlation Matrix for Consumer Sentiment Items: $10 vs. $5

	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	--	0.11	0.25	0.27	0.16
Item 2	0.10	--	0.42	0.39	0.12
Item 3	0.15	0.33	--	0.65	0.20
Item 4	0.17	0.25	0.60	--	0.19
Item 5	0.13	0.09	0.27	0.15	--

F2. COVARIANCE MATRICES

Table F.2A. Covariance Matrix for Patriotism Items: Incentive Condition

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 1	0.65									
Item 2	0.13	0.65								
Item 3	0.30	0.23	0.70							
Item 4	0.09	0.09	0.15	0.79						
Item 5	0.11	0.13	0.16	0.08	0.39					
Item 6	0.11	0.12	0.17	0.12	0.14	0.58				
Item 7	0.08	0.11	0.08	0.11	0.09	0.12	0.47			
Item 8	0.09	0.07	0.05	0.01	0.04	0.08	0.01	0.25		
Item 9	0.15	0.10	0.11	0.03	0.08	0.11	0.08	0.12	0.43	
Item 10	0.21	0.14	0.12	0.09	0.06	0.08	0.00	0.10	0.16	0.68

Table F.2B. Covariance Matrix for Patriotism Items: Control Condition

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 1	0.79									
Item 2	0.31	0.71								
Item 3	0.39	0.31	0.86							
Item 4	0.18	0.15	0.17	0.70						
Item 5	0.18	0.14	0.17	0.07	0.32					
Item 6	0.24	0.19	0.24	0.18	0.15	0.71				
Item 7	0.25	0.19	0.24	0.15	0.17	0.22	0.53			
Item 8	0.10	0.06	0.08	0.06	0.09	0.15	0.09	0.24		
Item 9	0.25	0.17	0.24	0.08	0.16	0.18	0.19	0.12	0.48	
Item 10	0.33	0.23	0.24	0.08	0.13	0.24	0.17	0.10	0.23	0.73

Table F.2C. Covariance Matrix for Conscientiousness Items: Incentive Condition

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 1	1.03									
Item 2	0.28	0.65								
Item 3	0.26	0.17	0.85							
Item 4	0.31	0.27	0.15	1.71						
Item 5	0.21	0.16	0.24	0.44	1.33					
Item 6	0.20	0.23	0.18	0.45	0.55	1.51				
Item 7	0.35	0.36	0.27	0.32	0.25	0.23	0.85			
Item 8	0.35	0.32	0.18	0.52	0.72	0.61	0.33	1.93		
Item 9	0.10	0.08	0.11	0.29	0.38	0.40	0.12	0.40	0.83	
Item 10	0.40	0.30	0.25	0.38	0.34	0.20	0.35	0.52	0.16	1.40

Table F.2D. Covariance Matrix for Conscientiousness Items: Control Condition

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 1	1.23									
Item 2	0.28	0.63								
Item 3	0.28	0.23	0.71							
Item 4	0.34	0.22	0.22	1.61						
Item 5	0.19	0.16	0.11	0.67	1.54					
Item 6	0.15	0.22	0.20	0.39	0.41	1.37				
Item 7	0.37	0.38	0.28	0.18	0.17	0.22	0.88			
Item 8	0.36	0.16	0.15	0.68	0.56	0.49	0.24	2.05		
Item 9	0.26	0.10	0.12	0.36	0.30	0.13	0.13	0.41	0.98	
Item 10	0.45	0.35	0.29	0.44	0.31	0.44	0.44	0.56	0.32	1.39

Table F.2E. Covariance Matrix for Impression Management Items: Incentive Condition

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 1	1.01									
Item 2	-0.11	0.87								
Item 3	0.20	0.08	0.83							
Item 4	0.23	0.03	0.16	1.06						
Item 5	0.06	0.13	0.15	0.22	1.24					
Item 6	0.14	0.07	0.30	0.15	0.21	0.78				
Item 7	-0.22	0.01	0.35	0.29	0.29	0.37	1.35			
Item 8	-0.07	0.15	0.15	0.05	0.27	0.11	0.09	0.77		
Item 9	0.20	0.02	0.23	0.25	0.26	0.39	0.29	0.10	0.95	
Item 10	0.09	0.09	0.19	0.22	0.19	0.17	0.20	0.14	0.18	0.80

Table F.2F. Covariance Matrix for Impression Management Items: Control Condition

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 1	0.93									
Item 2	-0.01	0.85								
Item 3	0.24	0.14	0.83							
Item 4	0.22	0.18	0.17	1.11						
Item 5	0.08	0.22	0.05	0.25	1.23					
Item 6	0.20	0.15	0.31	0.14	0.21	0.74				
Item 7	0.27	0.04	0.32	0.24	0.15	0.25	1.14			
Item 8	-0.11	0.15	0.02	0.04	0.21	0.10	-0.02	0.96		
Item 9	0.24	0.07	0.31	0.26	0.15	0.25	0.36	0.01	0.98	
Item 10	0.13	0.10	0.17	0.18	0.14	0.13	0.14	0.14	0.19	0.80

Table F.2G. Covariance Matrix for Consumer Sentiment Items: $10 Condition

	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	3.04				
Item 2	0.26	1.76			
Item 3	0.79	1.02	3.31		
Item 4	0.82	0.91	2.08	3.13	
Item 5	0.44	0.24	0.56	0.53	2.43

Table F.2H. Covariance Matrix for Consumer Sentiment Items: $5 Condition

	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	2.73				
Item 2	0.21	1.49			
Item 3	0.45	0.71	3.23		
Item 4	0.51	0.55	1.90	3.11	
Item 5	0.33	0.16	0.71	0.38	2.20

Table F.2I. Covariance Matrix for Consumer Sentiment Items: Control Condition

	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	2.86				
Item 2	0.47	1.75			
Item 3	1.35	0.65	3.53		
Item 4	1.33	0.68	1.76	3.31	
Item 5	0.27	-0.20	0.30	0.35	2.07

F3. RESULTS WITH IMPUTED VALUES

Table F.3A. Cronbach's Alpha Coefficient, by Incentive Condition

	Incentive Alpha	Control Alpha	*W*-statistic
Patriotism[1]	0.69 (n=523)	0.80 (n=374)	1.55 (p<0.001)
Conscientiousness[2]	0.77 (n=523)	0.76 (n=373)	1.04 (p=0.339)
Impression Management[3]	0.67 (n=446)	0.67 (n=564)	1.00 (p=0.499)
Consumer Sentiment			
$10 vs. Control	0.62 (n=297)	0.66 (n=232)	1.12 (p=0.183)
$5 vs. Control	0.60 (n=282)	0.66 (n=232)	1.18 (p=0.095)
$10 vs. $5	0.62 (n=297)	0.60 (n=282)	1.05 (p=0.339)

1. Values could not be imputed for one of the cases that skipped all ten of the patriotism items.

2. Values could not be imputed for two of the cases that skipped all ten of the conscientiousness items.

3. Values could not be imputed for six of the cases that skipped all ten of the impression management items.

Table F.3B. Differential Item Functioning: Patriotism (n=897)

	$\chi^2(1)$ **from Model 1**	$\chi^2(3)$ **from Model 2**	$\chi^2(2)$ **: Difference between Model 1 and Model 2**	***Pseudo* R^2: Difference between Model 1 and Model 2**
Way democracy works	482.23 (p<0.0001)	489.20 (p<0.0001)	6.97 (p=0.031)	0.005
Political influence in the world	406.88 (p<0.0001)	407.63 (p<0.0001)	0.75 (p=0.687)	0.001
Economic achievements	444.89 (p<0.0001)	445.10 (p<0.0001)	0.20 (p=0.905)	0.000
Social security system	201.09 (p<0.0001)	207.30 (p<0.0001)	6.21 (p=0.045)	0.006
Scientific / technological achievements	292.60 (p<0.0001)	294.40 (p<0.0001)	1.80 (p=0.407)	0.002
Achievements in sports	357.48 (p<0.0001)	358.69 (p<0.0001)	1.21 (p=0.546)	0.001
Achievements in arts and literature	292.02 (p<0.0001)	296.17 (p<0.0001)	4.15 (p=0.126)	0.003
Armed forces	140.14 (p<0.0001)	141.39 (p<0.0001)	1.25 (p=0.535)	0.001
History	355.36 (p<0.0001)	356.56 (p<0.0001)	1.20 (p=0.549)	0.001
Fair and equal treatment of all groups	339.37 (p<0.0001)	339.53 (p<0.0001)	0.16 (p=0.923)	0.000

Table F.3C. Differential Item Functioning: Conscientiousness (n=896)

	$\chi^2(1)$ **from Model 1**	$\chi^2(3)$ **from Model 2**	$\chi^2(2)$ **: Difference between Model 1 and Model 2**	***Pseudo* R^2: Difference between Model 1 and Model 2**
Always prepared	277.40 (p<0.0001)	274866 (p<0.0001)	1.26 (p=0.533)	0.001
Carry out plans	320.61 (p<0.0001)	324.65 (p<0.0001)	4.04 (p=0.133)	0.003
Pay attention to details	194.26 (p<0.0001)	196.22 (p<0.0001)	1.96 (p=0.375)	0.002
Waste time	403.68 (p<0.0001)	404.80 (p<0.0001)	1.12 (p=0.571)	0.001
Just enough work to get by	370.17 (p<0.0001)	378.16 (p<0.0001)	7.99 (p=0.018)	0.006
Don't see things through	357.52 (p<0.0001)	358.11 (p<0.0001)	0.59 (p=0.745)	0.001
Make plans and stick to them	310.96 (p<0.0001)	313.17 (p<0.0001)	2.21 (p=0.331)	0.002
Difficulty getting started doing work	4447.45 (p<0.0001)	449.34 (p<0.0001)	1.89 (p=0.389)	0.001
Avoid duties	269.47 (p<0.0001)	227.27 (p<0.0001)	2.80 (p=0.247)	0.002
Get chores done right away	396.64 (p<0.0001)	399.17 (p<0.0001)	2.53 (p=0.282)	0.002

Table F.3D. Differential Item Functioning: Impression Management (n=1010)

	$\chi^2(1)$ **from Model 1**	$\chi^2(3)$ **from Model 2**	$\chi^2(2)$ **: Difference between Model 1 and Model 2**	***Pseudo* R^2: Difference between Model 1 and Model 2**
Sometimes drive faster than speed limit	224.87 (*p*<0.0001)	229.50 (*p*<0.0001)	4.63 (*p*=0.099)	0.004
Never cover up mistakes	164.66 (*p*<0.0001)	168.64 (*p*<0.0001)	3.98 (*p*=0.137)	0.003
Sometimes tell lies if have to	394.97 (p<0.0001)	399.72 (*p*<0.0001)	4.75 (*p*=0.093)	0.003
Never swear	340.53 (*p*<0.0001)	341.91 (*p*<0.0001)	1.38 (*p*=0.502)	0.001
Never dropped litter on street	326.88 (p<0.0001)	330.97 (p<0.0001)	4.09 (*p*=0.129)	0.003
Occasions where taken advantage of someone	459.56 (p<0.0001)	460.35 (p<0.0001)	0.79 (*p*=0.674)	0.001
Taken leave from work/school even though not really sick	392.95 (p<0.0001)	399.13 (p<0.0001)	6.18 (*p*=0.046)	0.004
Never take things that don't belong to me	184.16 (p<0.0001)	185.40 (p<0.0001)	1.24 (*p*=0.538)	0.001
Said something bad about a friend behind his/her back	434.82 (p<0.0001)	434.83 (p<0.0001)	0.01 (*p*=0.995)	0.000
When hear people talking privately, avoid listening	282.35 (p<0.0001)	283.23 (p<0.0001)	0.88 (*p*=0.644)	0.001

Table F.3E. Differential Item Functioning: Consumer Sentiment: $10 vs. 0 (n=529)

	$\chi^2(1)$ **from Model 1**	$\chi^2(3)$ **from Model 2**	$\chi^2(2)$ **: Difference between Model 1 and Model 2**	***Pseudo R²*: Difference between Model 1 and Model 2**
Personal financial situation compared to one year ago	253.96 (p<0.0001)	258.94 (p<0.0001)	4.89 (p=0.087)	0.006
Personal financial situation compared to one year from now	153.41 (p<0.0001)	157.31 (p<0.0001)	3.90 (p=0.142)	0.006
Business conditions in the next 12 months	389.48 (p<0.0001)	391.82 (p<0.0001)	2.34 (p=0.310)	0.002
Good/bad times for the country in next five years	436.55 (p<0.0001)	436.68 (p<0.0001)	0.13 (p=0.937)	0.000
Good/bad time to buy major household items	107.35 (p<0.0001)	180.05 (p<0.0001)	0.70 (p=0.705)	0.001

Table F.3F. Differential Item Functioning: Consumer Sentiment: $5 vs. 0 (n=514)

	$\chi^2(1)$ **from Model 1**	$\chi^2(3)$ **from Model 2**	$\chi^2(2)$ **: Difference between Model 1 and Model 2**	***Pseudo R²*: Difference between Model 1 and Model 2**
Personal financial situation compared to one year ago	242.75 (p<0.0001)	249.41 (p<0.0001)	6.66 (p=0.036)	0.008
Personal financial situation compared to one year from now	136.74 (p<0.0001)	136.99 (p<0.0001)	0.25 (p=0.882)	0.002
Business conditions in the next 12 months	377.82 (p<0.0001)	378.60 (p<0.0001)	0.78 (p=0.677)	0.001
Good/bad times for the country in next five years	438.23 (p<0.0001)	442.39 (p<0.0001)	4.16 (p=0.125)	0.003
Good/bad time to buy major household items	103.29 (p<0.0001)	104.31 (p<0.0001)	1.02 (p=0.600)	0.002

Table F.3G. Differential Item Functioning: Consumer Sentiment: $10 vs. $5 (n=579)

	$\chi^2(1)$ **from Model 1**	$\chi^2(3)$ **from Model 2**	$\chi^2(2)$ **: Difference between Model 1 and Model 2**	***Pseudo* R^2: Difference between Model 1 and Model 2**
Personal financial situation compared to one year ago	202.44 (p<0.0001)	203.09 (p<0.0001)	0.65 (p=0.723)	0.001
Personal financial situation compared to one year from now	190.94 (p<0.0001)	192.15 (p<0.0001)	1.21 (p=0.546)	0.002
Business conditions in the next 12 months	403.15 (p<0.0001)	409.95 (p<0.0001)	6.80 (p=0.033)	0.003
Good/bad times for the country in next five years	479.93 (p<0.0001)	485.58 (p<0.0001)	5.65 (p=0.059)	0.005
Good/bad time to buy major household items	125.64 (p<0.0001)	126.24 (p<0.0001)	0.60 (p=0.741)	0.001

Table F.3H. Fit Indices for CFA Models, by Incentive Condition: Patriotism

	Chi-square (df, *p*-value)	**RMSEA (90% CI)**	**SRMR**	**CFI**	**Diff Btwn Chi-square values**
Initial Models					
Incentive (n=523)	126.39 (df=35, p<0.01)	0.07 (0.06, 0.08)	0.05	0.90	--
Control (n=374)	91.29 (df=35, p<0.01)	0.07 (0.05, 0.09)	0.05	0.96	--
Configural Invariance Model (n=897)	217.67 (df=70, p<0.01)	0.07 (0.06, 0.08)	0.05	0.94	--
Metric Invariance Model (n=897)	229.33 (df=79, p<0.01)	0.07 (0.06, 0.08)	0.06	0.94	11.66 (df=9, p=0.23)

Table F.3I. Fit Indices for CFA Models, by Incentive Condition: Conscientiousness

	Chi-square (df, *p*-value)	**RMSEA (90% CI)**	**SRMR**	**CFI**	**Diff Btwn Chi-square values**
Initial Models					
Incentive (n=523)	194.40 (df=35, $p<0.01$)	0.11 (0.10, 0.12)	0.07	0.90	--
Control (n=373)	178.33 (df=35, $p<0.01$)	0.12 (0.11, 0.14)	0.08	0.87	--
Configural Invariance Model (n=896)	372.73 (df=70, $p<0.01$)	0.11 (0.10, 0.12)	0.08	0.89	--
Metric Invariance Model (n=896)	375.29 (df=79, $p<0.01$)	0.11 (0.10, 0.12)	0.08	0.89	2.56 (df=9, $p=0.98$)

Table F.3J. Fit Indices for CFA Models, by Incentive Condition: Impression Management

	Chi-square (df, *p*-value)	**RMSEA (90% CI)**	**SRMR**	**CFI**	**Diff Btwn Chi-square values**
Initial Models					
Incentive (n=446)	110.58 (df=35, $p<0.01$)	0.07 (0.06, 0.09)	0.06	0.90	--
Control (n=564)	158.03 (df=35, $p<0.01$)	0.08 (0.07, 0.10)	0.07	0.87	--
Configural Invariance Model (n=1010)	286.61 (df=70, $p<0.01$)	0.08 (0.07, 0.09)	0.07	0.88	--
Metric Invariance Model (n=1010)	280.44 (df=79, $p<0.01$)	0.07 (0.07, 0.08)	0.07	0.88	6.17 (df=9, $p=0.73$)

Table F.3K. Fit Indices for CFA Models, by Incentive Condition: Consumer Sentiment

	Chi-square (df, *p*-value)	RMSEA (90% CI)	SRMR	CFI	Diff Btwn Chi-square values
Initial Models					
$10 (n=297)	3.52 (df=5, $p<0.62$)	0.00 (0.00, 0.07)	0.03	1.00	--
$5 (n=282)	7.95 (df=5, $p<0.19$)	0.04 (0.00, 0.10)	0.03	0.99	
Control (n=232)	6.45 (df=5, $p<0.26$)	0.03 (0.00, 0.10)	0.04	0.99	--
Configural Invariance Models					
$10 vs. Control	9.98 (df=10, $p=0.44$)	0.00 (0.00, 0.07)	0.04	1.00	--
$5 vs. Control	14.40 (df=10, $p=0.16$)	0.04 (0.00, 0.08)	0.08	0.99	--
$10 vs. $5	11.47 (df=10, $p=0.32$)	0.02 (0.00, 0.07)	0.03	1.00	--
Metric Invariance Model					
$10 vs. Control	23.88 (df=14, $p=0.05$)	0.05 (0.00, 0.09)	0.07	0.98	13.90 (df=4, $p=0.01$)
$5 vs. Control	31.13 (df=14, $p<0.01$)	0.07 (0.04, 0.10)	0.08	0.95	16.73 (df=4, $p<0.01$)
$10 vs. $5	13.73 (df=14, $p=0.47$)	0.00 (0.00, 0.05)	0.04	1.00	2.26 (df=4, $p=0.69$)

Table F.3L. Modification Indices Produced by Metric Invariance Models: Consumer Sentiment

	$10 vs. Control	$5 vs. Control
Personal financial situation compared to one year ago	12.51**	15.65**
Personal financial situation compared to one year from now	1.48	0.21
Business conditions in the next 12 months	0.29	0.61
Good/bad times for the country in next five years	1.70	2.47
Good/bad time to buy major household items	0.02	0.23

* $p<0.05$; ** $p<0.01$

Table F.3M. Fit Indices for Metric Invariance Models with One Equality Constraint Removed: Consumer Sentiment

	Chi-square (df, *p*-value)	**RMSEA (90% CI)**	**SRMR**	**CFI**	**Diff Btwn Chi-square values**
$10 Incentive vs. Control	11.21 (df=13, *p*=0.59)	0.00 (0.00, 0.05)	0.04	1.00	1.23 (df=3, *p*=0.746)
$5 Incentive vs. Control	15.38 (df=13, *p*=0.28)	0.02 (0.00, 0.07)	0.04	0.99	0.98 (df=3, *p*=0.806)

Table F.3N. Modification Indices After Releasing One Equality Constraint: Consumer Sentiment

	$10 vs. Control	**$5 vs. Control**
Personal financial situation compared to one year ago	--	--
Personal financial situation compared to one year from now	0.82	0.01
Business conditions in the next 12 months	0.37	0.59
Good/bad times for the country in next five years	0.05	0.43
Good/bad time to buy major household items	0.17	0.03

* $p<0.05$; ** $p<0.01$

REFERENCES

Armstrong, J. S. (1975). Monetary incentives in mail surveys. *Public Opinion Quarterly, 39,* 111-116.

Atrostic, B. K., Bates, N., Burt, G., Silberstein, A. (2001). Nonresponse in U.S. government household surveys: Consistent measures, recent trends, and new insights. *Journal of Official Statistics, 17,* 209-226.

Bates, N., Dalhamer, J., & Singer, E. (2008). Privacy concerns, too busy, or just not interested: Using doorstep concerns to predict survey nonresponse. *Journal of Official Statistics, 24,* 591–612.

Battaglia, M. P., Khare, M., Frankel, M. R., Murray, M. C., Buckley, P., & Peritz, S, (2008). Response rates: How have they changed and where are they headed? In J. M. Lepkowski, C. Tucker, J. M. Brick, E. de Leeuw, L. Japec, P. Lavrakas, M. W. Link, & R. L. Sangster. (Eds.), *Advances in telephone survey methodology*. (pp. 529–560). New York: Wiley.

Baumgartner, R., & Rathbun, P. (1996). *Prepaid monetary incentives and mail survey response rates*. Paper presented at the annual conference of the American Statistical Association. Chicago, IL.

Bradburn, N. M., Sudman, S., & Associates. (1979). *Improving interview methods and questionnaire design.* San Francisco: Jossey-Bass.

Belli, R. F., Traugott, M. W., & Beckmann, M. N. (2001). What leads to voting Overreports? Contrasts of overreporters to validated voters and admitted nonvoters in the American National Election Studies. *Journal of Official Statistics, 17,* 479-498.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238-246.

Berk, M. L., Mathiowetz, E. P., Ward, E. P, & White, A. A. (1987). The effect of prepaid and promised incentives: Results of a controlled experiment. *Journal of Official Statistics, 3,* 449-457.

Berlin, M., Mohadjer, L., Waksberg, J., Kolstad, A., Kirsch, D., Rock, D. & Yamamoto, K. (1992). *An experiment in monetary incentives*. Proceedings of the Survey Research Methods Section of American Statistical Association.

Biner, P. M., & Kidd, H. J. (1994). The interactive effects of monetary incentive justification and questionnaire length on mail survey response rates. *Psychology and Marketing, 11,* 483-492.

Blau, P. M. (1964). *Exchange and power in social life*. New York: Wiley.

Blumberg, S. J., & Luke, J. V. (2007). Coverage bias in traditional telephone surveys of low-income and young adults. *Public Opinion Quarterly, 71*, 734-749.

Bohrnstedt, G. W. (2010). Measurement models for survey research. In P. V. Marsden, & J. W. Wright. (Eds.), *Handbook of Survey Research*. (pp. 347-404). United Kingdom: Emerald.

Bolkan, S., Goodboy, A. K., & Griffin, D. J. (2011). Teacher leadership and intellectual stimulation: Improving students' approaches to studying through intrinsic motivation. *Communication Research Reports, 28*, 337-346.

Brehm, J. (1994). Stubbing our toes for a foot in the door? Prior contact, incentives and survey response. *International Journal of Public Opinion Research, 6*, 45-63.

Brehm, J. (1993). *The phantom respondents: Opinion surveys and political representation*. Ann Arbor: University of Michigan Press.

Brick, J. M., Montaquila, J., Hagedorn, M. C., Roth, S. B., & Chapman, C. (2005). Implications for RDD design from an incentive experiment. *Journal of Official Statistics, 21*, 571-589.

Brick, J. M., & Willliams, D. (forthcoming). Reasons for increasing nonresponse in U.S. household surveys. In D. S. Massey, & R. Tourangeau (Eds.*), The ANNALS of the American Academy of Political and Social Science (Special Issue: The Non-Response Challenge to Measurement in Social Science)*, Volume 645.

Buchanan, T., Johnson, J. A., & Goldberg, L. R. (2005). Implementing a five-factor personality inventory for use on the Internet. *European Journal of Psychological Assessment, 21*, 116-128.

Byrne, B. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming.* New Jersey: Erlbaum.

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty, 19*, 7-42.

Cannell, C. F., & Fowler, F. J. (1963). Comparison of a self-enumerative procedure and a personal interview: A validity study. *Public Opinion Quarterly, 27*, 250-264.

Cantor, D., O'Hare, B. C., & O'Connor, K. S. (2008). The use of monetary incentives to reduce nonresponse in random digit dial telephone surveys. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. de Leeuw, L. Japec, P. Lavrakas, M. W. Link, & R. L. Sangster. (Eds.), *Advances in telephone survey methodology*. (pp. 471–498). New York: Wiley.

Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing versus the Internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly, 73,* 641-678.

Church, A. H. (1993). Estimating the effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly, 57,* 62-79.

Cole, S. R. (1999). Assessment of differential item functioning in the Perceived Stress Scale-10. *Journal of Epidemiology and Community Health, 53,* 319-320.

Cole, M. S., Bedeian, A. G., & Feild, H. S. (2006). The measurement equivalence of web-based and paper-and-pencil measures of transformational leadership: A multinational test. *Organizational Research Methods, 9,* 339-368.

Conrad, F. G., Brown, N. R., & Cashman, E. R. (1998). Strategies for estimating behavioral frequency in survey interviews. *Memory, 6,* 339-366.

Coogan, P. F., & Rosenberg, L. (2004). Impact of a financial incentive on case and control participation in a telephone interview. *American Journal of Epidemiology, 160,* 295-298.

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO personality inventory. *Psychological Assessment, 4,* 5-13.

Couper, M. P., Singer, E., Conrad, F. G., & Groves, R. M. (2008). Risk of disclosure, perceptions of risk, and concerns about privacy and confidentiality as factors in survey participation. *Journal of Official Statistics, 24,* 255-275.

Crane, P. K., Gibbons, L. E., Ocepek-Welikson, K., Cook, K., Cella, D., Narasimhalu, K., Hays, R. D., Teresi, J. A. (2007). A comparison of three sets of criteria for determining the presence of differential item functioning using ordinal logistic regression. *Quality of Life Research, 16,*69-84.

Creighton, K. P., King, K. E., & Martin, E. A. (2007). *The use of monetary incentives in Census Bureau longitudinal surveys*. In Research Report Series (Survey Methodology 2007-2). Washington DC: U.S. Census Bureau.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334.

Crowne, D., & Marlowe, D. (1964). *The approval motive.* New York: Wiley.

Curtin, R., Presser, S., & Singer, E. (2005). Changes in telephone nonresponse over the past quarter century. *Public Opinion Quarterly, 69*, 87-98.

Curtin, R., Singer, E., & Presser, S. (2007). Incentives in random digit dial telephone surveys: A replication and extension. *Journal of Official Statistics, 23*, 91-105.

Davern, M., Rockwell, T. H., Sherrod, R., & Campbell, S. (2003). Prepaid monetary incentives and data quality in face-to-face interviews. *Public Opinion Quarterly, 67*, 139-147.

Davidov, E., Schmidt, P., & Schwartz, S. H. (2008). Bringing values back in: The adequacy of the European Social Survey to measure values in 20 countries. *Public Opinion Quarterly, 72*, 420-445.

De Leeuw, E., & de Heer, W. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge & R. Little. (Eds.), *Survey Nonresponse*. (pp. 41–54). New York: Wiley.

Diamantopoulos, A., & Siguaw, J. (2000). *Introducing LISREL: A Guide for the Uninitiated*. London: SAGE.

Dillman, D. A. (1996). *Token financial incentives and the reduction of nonresponse error in mail surveys*. Proceedings of the Government Statistics Section of the American Statistical Association.

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.

Dirmaier, J., Harfst, T., Koch, U., & Schulz, H. (2007). Incentives increased return rates but did not influence partial nonresponse or treatment outcome in a randomized trial. *Journal of Clinical Epidemiology, 60*, 1263-1270.

Durvasula, S., Lysonski, S., & Watson, J. (2001). Does vanity describe other cultures?: a cross-cultural examination of the vanity scale. *Journal of Consumer Affairs, 35*, 180-199.

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R., & Kwan, I. (2002). Increasing response rates to postal questionnaires: Systematic review. *British Medical Journal, 324*, 1183-1191.

Eyerman, J., Bowman, K., Butler, D., & Wright, D. (2005). The differential impact of incentives on refusals: Results from the 2001 National Household Survey on Drug Abuse incentive experiment. *Journal of Economic & Social Measurement, 30*, 157-169.

Feldt, L. S. (1969). A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests. *Psychometrika, 34,* 363-373.

Feldt, L. S., & Kim, S. (2006). Testing the difference between two alpha coefficients with small samples of subjects and raters. *Educational and Psychological Measurement, 66,* 589-600.

Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row, Peterson.

Fox, R. J., Crask, M. R., Kim, J. (1988). Mail survey response rate: A meta-analysis of selected techniques for inducing response. *Public Opinion Quarterly, 52,* 467-491.

Fransson, A. (1977). On qualitative differences in learning: IV – Effects of intrinsic motivation and extrinsic test anxiety on process and outcome. *British Journal of Educational Psychology, 47,* 244-257.

Fricker, S., & Tourangeau, R. (2010). Examining the relationship between nonresponse propensity and data quality in two national household surveys. *Public Opinion Quarterly, 74,* 934-955.

Friedman, E. M., Clusen, N. A., & Hartzell, M. (2003). *Better late? Characteristics of late respondents to a health care survey.* Proceedings of the Survey Research Methods Section of American Statistical Association.

Furse, D. H., & Stewart, D. W. (1982). Monetary incentives versus promised contribution to charity: New evidence on mail survey response, *Journal of Marketing Research, 19,* 375-380.

Gelin, M. A., & Zumbo, B. D. (2003). Differential item functioning results may change depending on how an item is scored: An illustration with the center for epidemiologic studies depression scale. *Educational and Psychological Measurement, 63,* 65-74.

Gelman, A., Stevens, M., & Chan, V. (2003). Regression modeling and meta-analysis for decision making: A cost-benefit analysis of incentives in telephone surveys. Journal of *Business and Economic Statistics, 21,* 213-225.

Godwin, R. K. (1979). The consequences of large monetary incentives in mail surveys of elites. *Public Opinion Quarterly, 43,* 378-387.

Goetz, E. G., Tyler, T. R., & Cook, F.L. (1984). Promised incentives in media research: A look at data quality, sample representativeness, and response rate. *Journal of Marketing Research, 21,* 148-54.

Goldenberg, K. L., McGrath D. E., and Tan, L. (2009). *The effects of incentives on the Consumer Expenditure Interview Survey*. Paper presented at the Annual Meeting of the American Association for Public Opinion Research. Hollywood, FL.

Goodstadt, M. S., Chung, L., Kronitz, R., & Cook, G. (1977). Mail survey response rates: Their manipulation and impact. *Journal of Marketing Research, 14*, 391-395.

Göritz, A. S. (2006). Incentives in web studies: Methodological issues and a review. International *Journal of Internet Science, 1*, 58-70.

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review, 25*, 161–178.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.

Groves, R. M. (2006). Nonresponse rates and nonresponse error in household surveys. *Public Opinion Quarterly, 70*, 646–675.

Groves, R. M. (2008, March). *Survey respondent incentives: Research and practice*. Speech presented at Federal Committee on Statistical Methodology, Washington, D.C.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: Wiley.

Groves, R. M., Couper, M. P., Presser, S., Singer, E., Tourangeau, R., Acosta, G. P., & Nelson, L. (2006). Experiments in producing nonresponse bias. *Public Opinion Quarterly, 70*, 720-736.

Groves, R. M., Dillman, D., Eltinge, J., & Little, R. (2002). *Survey Nonresponse*. New York: Wiley.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly, 72*, 167-189.

Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly, 68*, 2-31.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation. *Public Opinion Quarterly, 64*, 299-308.

Hall, M. (1995). Patient satisfaction or acquiescence: Comparing mail and telephone survey results. *Journal of Health Care Marketing, 15*, 54-61.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. California: Sage.

Hansen, R. A. (1980). A self-perception interpretation of the effect of monetary and nonmonetary incentives on mail survey respondent behavior. *Journal of Marketing Research, 17,* 77-83.

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly, 67,* 79-125.

Holbrook, A. L., Krosnick, J. A., & Pfent, A. (2008). The causes and consequences of response rates in surveys by the news media and government contractor survey research firms. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. de Leeuw, L. Japec, P. Lavrakas, M. W. Link, & R. L. Sangster. (Eds.), *Advances in telephone survey methodology*. (pp. 471–498). New York: Wiley.

Homans, G. C. (1961). *Social behavior: Its elementary forms*. New York: Harcourt Brace.

Hopkins, K. D., & Gullickson, A. R. (1992). Response rates in survey research: A meta-analysis of the effects of monetary gratuities. *The Journal of Experimental Education, 61,* 52-62.

Hu, L, & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76-99). London: Sage.

Hu , L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives, *Structural Equation Modeling, 6,* 1-55.

James, J. M., & Bolstein, R. (1990). The effect of monetary incentives and follow-up mailings on the response rate and response quality in mail surveys. *Public Opinion Quarterly, 54,* 346-361.

James, J. M., & Bolstein, R. (1992). Large monetary incentives and their effect on mail survey response rates. *Public Opinion Quarterly, 56,* 442-453.

Jensen, P. S., Watanabe, H. K., Richter, J. E. (1999). Who's up first? Testing for order effects in structured interviews using a counterbalanced experimental design. *Journal of Abnormal Child Psychology, 27,* 439-445.

Jöreskog, K.G. & Sörbom, D. (2006). *LISREL 8.80 for Windows [Computer Software]*. Lincolnwood, IL: Scientific Software International, Inc.

Juster, T., & Smith, J. P. (1997). Improving the quality of economic data: Lessons from the HRS and AHEAD. *Journal of the American Statistical Association, 92,* 1268–1278.

Kaminska, O., McCutcheon, A., & Billiet, J. (2010). Satisficing among reluctant respondents in a cross-national context. *Public Opinion Quarterly, 74,* 956-984.

Keeter, S., Kennedy, C., Clark, A., Tompson, T., & Mokrzycki, M. (2007). What's missing from national landline RDD surveys? The impact of the growing cell-only population. *Public Opinion Quarterly, 71,* 772-792.

Kennedy, C., & Everett, S. E. (2011). Use of cognitive shortcuts in landline and cell phone surveys. *Public Opinion Quarterly, 75*, 336–348.

Kleinjans, K.J. (2008). *The man of the house: How the use of household head characteristics may lead to omitted variable bias*. Unpublished paper.

Kline, R. B. (2011). *Principles and practices of structural equation modeling.* New York: Guilford.

Krenzke, T., Mohadjer, L., Ritter, G., & Gadzuk, A. (2005). Incentive effects on self-report of drug use and other measures of response quality in the Alcohol and Drug Services study. *Journal of Economic and Social Measurement, 30,* 191-217.

Kreuter, F., McCulloch, S., Presser, S., & Tourangeau, R. (2011). The effects of asking filter questions in interleafed versus grouped format. *Sociological Methods & Research, 40,* 88-104.

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly, 72*, 847-865.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5,* 213-236.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50,* 537-567.

Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly, 51*, 201-219.

Krosnick, J. A., Narayan, S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. *New Directions for Evaluation, 70*, 29-44.

Kulka, R. A., Eyerman, J., & McNeeley, M. E. (2005). The use of monetary incentives in federal surveys on substance use and abuse. *Journal of Economic and Social Measurement, 30,* 233-249.

Levine, D. W., Kaplan, R. M., Kripke, D. F., Bowen, D. J., Naughton, M. J., & Shumaker, S. A. (2003). Factor structure and measurement invariance of the Women's Health Initiative Insomnia Rating Scale. *Psychological Assessment, 15,* 123-136.

Link, M., & Mokdad, A. H. (2005). Advance letters as a means of improving respondent cooperation in RDD studies: A multi-state experiment. *Public Opinion Quarterly, 69,* 572-587.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. New Jersey: Erlbaum.

Luman, E. T., Ryman, T. K., & Sablan, M. (2009). Estimating vaccination coverage: Validity of household-retained vaccination cards and parental recall. *Vaccine, 27,* 2534-2539.

Lucas, C. P., Fisher, P., Piacentini, J., Zhang, H., Jensen, P. S., Shaffer, D., Dulcan, M., Schwab-Stone, M., Regier, D., & Canino, G. (1999). Features of interview questions associated with attenuation of symptom reports. *Journal of Abnormal Child Psychology, 27,* 429-437.

Lynn, P., & Sturgis, P. (1997). *Boosting survey response through a monetary incentive and fieldwork procedures: An experiment*. Survey Methods Centre Newsletter, 17, 18-22.

Mack, S., Huggins, V., Keathley, D., & Sundukchi, M. (1998). *Do monetary incentives improve response rates in the survey of income and program participation?* Paper presented at the annual conference of the American Statistical Association. Dallas, TX.

Magura, S., & Kang, S.-Y. (1997). *The validity of self-reported drug use: Improving the accuracy of survey estimates* (pp. 227-246). Rockville, MD: National Institute on Drug Abuse.

Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public Opinion Quarterly. 72,* 914-934.

McDaniel, S. W., & Rao, C. P. (1980). The effect of monetary inducement on mailed questionnaire response quality. *Journal of Marketing Research, 17,* 265-268.

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology, 4,* 103-20.

Meade, A. W. & Lautenschlager, G. J. (2004). A comparison of IRT and CFA methodologies for establishing measurement equivalence. *Organizational Research Methods, 7,* 361-388.

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research, 13,* 127-143.

Medway, R., Tourangeau, R., Viera, L., Turner, S., & Marsh, S. (2011). *Exploring the impact of incentives on multiple indicators of data quality.* Paper presented at the annual conference of the American Association for Public Opinion Research. Phoenix, AZ.

Miller, K. J. (1996). *The influence of different techniques on response rates and nonresponse error in mail surveys*. (Unpublished master's thesis). Western Washington University, Bellingham MA.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance.* New York: Routledge.

Mizes, J. S., Fleece, E. L., & Roos, C. (1984). Incentives for increasing return rates: Magnitude levels, response bias, and format. *Public Opinion Quarterly, 48,* 794-800.

Moore, J. C., Stinson, L. L., & Welniak, Jr. E. J. (1999). Income reporting in surveys: Cognitive issues and measurement error. In M. G. Sirken, D. J. Herrmann, S. Schechter, N. Schwarz, J. M. Tanur, & R. Tourangeau (Eds.), *Cognition and survey research* (pp. 155–173). New York: Wiley.

Moyer, A., & Brown, M. (2008). Effect of participation incentives on the composition of national health surveys. *Journal of Health Psychology, 13,* 870-873.

Narayanan, P., & Swaminathan, H. (1996). Identification of items that show nonuniform DIF. *Applied Psychological Measurement, 20,* 257-274.

Nederhof, A. J. (1983). The effects of material incentives in mail surveys: Two studies. *Public Opinion Quarterly, 47,* 103-111.

Osterlind, H .T., & Everson, S. J. (2009). *Differential item functioning*. California: Sage.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Personality Processes and Individual Differences, 46,* 598-609.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 51-76). New Jersey: Erlbaum.

Peck, J. K., & Dresch, S. P. (1981). Financial incentives, survey response and sample representativeness: Does money matter? *Review of Public Use Data, 9,* 245-266.

Petrolia, D. R., & Bhattacharjee, S. (2009). Revisiting incentive effects: evidence from a random-sample mail survey on consumer preferences for fuel ethanol. *Public Opinion Quarterly, 73,* 537-550.

Presser, S., & Stinson, L. (1998). Data collection mode and social desirability bias in self-reported religious attendance. *American Sociological Review, 63,* 137-145.

Raghunathan, T.E., Solenberger, P. W., & Van Hoewyk, J. (2002). *IVEware: Imputation and variance estimation software user guide.* Ann Arbor, MI: Survey Research Center, Institute for Social Research.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114,* 552-566.

Roberts, C. E., Jäckle, A., & Lynn, P. (2007). *Causes of mode effects: Separating out interviewer and stimulus effects in comparisons of face-to-face and telephone surveys*. Proceedings of the Social Research Methods Section of the American Statistical Association.

Rush, M. C., Phillips, J. S., & Panel, P. E. (1978). Subject recruitment bias: The paid volunteer subject. *Perceptual and Motor Skills, 47,* 443-449.

Ryu, E., Couper, M. P., & Marans, R. W. (2005). Survey incentives: Cash vs. in-kind; Face-to-face vs. mail; Response rates vs. nonresponse error. *International Journal of Public Opinion Research, 18,* 89-106.

Scandura, T. A., Williams, E. A., & Hamilton, B.A. (2001). Perceptions of organizational politics in the United States and the Middle East: Measuring invariance across two cultures. In C.A. Schriesheim & L.L. Neider (Eds.), *Research in Management, Vol. 1* (pp. 99-130). Greenwich, CT: Information Age Publishing.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18,* 210-222.

Shettle, C., & Mooney, G. (1999). Monetary incentives in U.S. government surveys. *Journal of Official Statistics, 15,* 231-250.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63,* 129-138.

Singer, E., Van Hoewyk, J., Gebler, N., Raghunathan, T., & McGonagle, K. (1999). The effect of incentives on response rates in interviewer-mediated surveys. *Journal of Official Statistics, 15,* 217-230.

Singer, E., Van Hoewyk, J., & Maher, M. P. (2000). Experiments with incentives in telephone surveys. *Public Opinion Quarterly, 64,* 171-188.

Singer, E., & Ye, C. (forthcoming). The use and effects of incentives in surveys. In D. S. Massey, & R. Tourangeau (Eds.*), The ANNALS of the American Academy of Political and Social Science (Special Issue: The Non-Response Challenge to Measurement in Social Science),* Volume 645.

Smith, T. W. (1995). Trends in non-response rates. *International Journal of Public Opinion Research, 7,* 157-171.

Steeh, C., Kirgis, N., Cannon, B., & DeWitt, J. (2001). Are they really as bad as they seem? Nonresponse rates at the end of the twentieth century. *Journal of Official Statistics, 21,* 227-247.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42,* 893-98.

Steiger J. H., & Lind, J. C. (1980, May). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society; Iowa, City, IA.

Storms, V., & Loosveldt, G. (2004). Who responds to incentives? *Field Methods, 16,* 414-421.

Sudman, S., & Bradburn, N. (1974). *Response effects in surveys: A review and synthesis*. Chicago: Aldine.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27,* 361-370.

Tourangeau, R., Groves, R. M., & Redline, C. D. (2010). Sensitive topics and reluctant respondents: Demonstrating a link between nonresponse bias and measurement error. *Public Opinion Quarterly, 74,* 413-432.

Tourangeau, R., Rasinski, K., Jobe, J. B., Smith, T. W., & Pratt, W. (1997). Sources of error in a survey of sexual behavior. *Journal of Official Statistics, 13,* 1-18.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response.* New York: Cambridge.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133,* 859-883.

Traugott, M. W., & Katosh, J. P. (1979). Response validity in surveys of voting behavior. *Public Opinion Quarterly, 43,* 359-377.

Trice, A. D. (1984). Hotel ratings IV: Recruitment and incentives. *Psychological Report, 55,* 58.

Triplett, T., Blair, J., Hamilton, T., & Kang, Y. C. (1996). *Initial cooperators vs. converted refusers: Are there response behavior differences?* Proceedings of the Survey Research Methods Section of American Statistical Association.

Trussell, N., & Lavrakas, P. J. (2004). The influence of incremental increases in token cash incentives on mail survey response: Is there an optimal amount? *Public Opinion Quarterly, 68,* 349-367.

Tzamourani, P., & Lynn, P. (1999). *The effect of monetary incentives on data quality – Results from the British Social Attitudes Survey 1998 experiment*. Centre for Research into Elections and Social Trends Working Paper.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3,* 4-69.

Warriner, K., Goyder, J., Gjertsen, H., Hohner, P., & McSpurren, K. (1996). Charities, no; Lotteries, no; Cash, yes. *Public Opinion Quarterly, 60,* 542-562.

Welkenhuysen-Gybels, J., Billiet, J., & Cambré, B. (2003). Adjustment for acquiescence in the assessment of the construct equivalence of Likert-type score items. *Journal of Cross-Cultural Psychology, 34,* 702-722.

Wheeler, W. J., Lazo, J. K., Heberling, M. T., Fisher, A. N., & Epp, D. J. (1997). *Monetary incentive response effects in contingent valuation mail surveys*. Draft paper. Penn State University.

Whitaker, B., & McKinney, J. (2007). Assessing the measurement invariance of latent job satisfaction ratings across survey administration modes for respondent subgroups: A MIMIC modeling approach. *Behavior Research Methods, 39,* 502-509.

Willimack, D. K., Schuman, H., Pennell, B., & Lepkowski, J. M. (1995). Effects of a prepaid nonmonetary incentive on response rates and response quality in a face-to-face survey. *Public Opinion Quarterly, 59,* 78-92.

Yammarino, F. J., Skinner, S. J., & Childers, T. L. (1991). Understanding mail survey response behavior: A meta-analysis. *Public Opinion Quarterly, 55,* 613-639.

Yan, T., Tourangeau, R., & Arens, Z. (2004). *When less is more: Are reluctant respondents poor reporters?* Paper presented at the Annual Meetings of the American Statistical Association. Toronto, Ontario.

Yu, J., & Cooper, H. (1983). A quantitative review of research design effects on response rates to questionnaires. *Journal of Marketing Research, 20,* 36-44.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores.* Ottawa ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Zumbo, B. D., & Thomas, D. R. (1997) *A measure of effect size for a model-based approach for studying DIF*. Working Paper of the Edgeworth Laboratory for Quantitative Behavioral Science, University of Northern British Columbia: Prince George, B.C.