Published in final edited form as:

Eval Rev. 2009 April ; 33(2): 138–158. doi:10.1177/0193841X08326463.

Use of Web and In-Person Survey Modes to Gather Data From Young Adults on Sex and Drug Use: An Evaluation of Cost, Time, and Survey Error Based on a Randomized Mixed-Mode Design

Barbara J. McMorris, Renee S. Petrie, Richard F. Catalano, Charles B. Fleming, Kevin P. Haggerty, and Robert D. Abbott

Social Research Development Group School of Social Work University of Washington 9725 3rd Avenue NE, Suite 401 Seattle, WA 98115

Abstract

In a randomized test of mixed-mode data collection strategies, 386 participants in the Raising Healthy Children (RHC) Project were either (1) asked to complete a survey over the Internet and later offered the opportunity to complete the survey in person or (2) first offered an in-person survey, with Web follow-up. The web-first condition resulted in cost savings while the overall completion rates for the two conditions were similar. On average, in-person-first condition participants completed surveys earlier in the field period than web-first condition participants. Based on intent-to-treat analyses, little evidence of condition effects on response bias, with respect to rates or levels of reported behavior, was found.

Keywords

Web survey; mixed-mode; longitudinal panel; substance use; sexual behavior; sensitive behavior

Introduction

Researchers who gather survey data to study complicated and sensitive health behaviors face the challenge of designing affordable research strategies that produce timely, valid, and reliable data, particularly as response rates to surveys have steadily declined over the past two decades (Krosnick 1999; Steeh 1981; Singer 2006). When gathering information among highly mobile populations, such as young adults as they leave high school, these challenges are magnified. One crucial decision in study design is choice of survey mode. Researchers must balance the practical constraints of time and cost against the sampling, coverage, nonresponse, and measurement error associated with each mode (Weisberg 2005; Groves 2004).

Each of the most common survey modes has relative advantages and disadvantages. Mail-out surveys save costs of interviewer time and travel but result in low completion rates, particularly among some portions of the population, and higher rates of incomplete data on individual items (Dillman 2000). Telephone surveys commonly save costs on interviewer travel times, allow for shorter survey field periods, and provide higher completion rates and less nonresponse bias. However, coverage problems arise due to call screening and cell-phone-only households (Dillman 2000; Groves 2004), and this survey mode may produce measurement error when respondents provide sensitive information over the telephone, resulting in social desirability bias (de Leeuw 2005; Groves 2004; Kraus and Augustin 2001; Parks, Pardi, and Bradizza 2006). In-person interviews are often the benchmark against which other modes are compared because of low nonresponse and coverage error, as well as low rates of incomplete data. Also, social desirability effects may be reduced by making portions of an interview with sensitive

questions self-administered, with respondents recording answers themselves, either on paper or into a laptop computer. However, in-person interviews are time consuming, often requiring relatively long field periods, increasing expense due to interviewer time and travel costs (Groves 2004).

Recently, the use of World Wide Web data collection methods has become more common (Dillman 2000; Couper 2000; Kypri, Gallagher, and Cashell-Smith 2004). Compared to other survey modes, web surveys are less expensive and faster to implement (Clayton and Werking 1998; Leece 2004; Schleyer and Forrest 2000; Shannon and Bradshaw 2002). Web surveys face challenges with regard to coverage and nonresponse error due to less than universal access to the internet. Some subgroups of potential respondents are less able or willing to complete surveys over the internet. For instance, there is some evidence that males are less likely than females to complete surveys via the Web (Weisberg 2005; Groves and Couper 1998; Kwak and Radler 2002; McCabe et al. 2002). Also, web surveys are prone to incomplete data due to the absence of an interviewer or proctor to monitor a participant's progress through the interview, with respondents often breaking off interviews in the middle (Couper 2000). Yet, although many early web surveys have been plagued by low response rates, researchers have achieved impressive results with this method when using evidence-based approaches with populations who have high levels of access to and regular use of the Internet (Kypri, Gallagher, and Cashell-Smith 2004; McCabe et al. 2002).

The relative advantages and disadvantages of different survey modes have resulted in the desire to use mixed-mode survey designs to maximize the relative advantage and minimize the disadvantages of the different approaches. Indeed, the methodological literature focusing on mixed-mode strategies is increasing (see Dillman and Christian 2005; Groves 2004; de Leeuw 2005; Weisberg 2005). Mixed-mode strategies offer the ability to capitalize on the strengths of different survey modes while neutralizing their weaknesses. A common strategy for realizing these benefits is adopting mixed-mode designs that begin with the less expensive mode to complete the maximum number of interviews at minimal cost and follow with the more expensive modes to improve completion rates (Groves 2004).

An important consideration in mixed-mode strategies is the mode impact on measurement error. In cases where a study variable of interest is associated with the mode completed, this can result in error being nonrandom across subjects depending on mode completed.

Measurement differences that have been found are mainly attributed to the presence or absence of an interviewer, question format, and channel of communication (oral or written) (Kraus and Augustin 2001; Parks, Pardi, and Bradizza 2006; de Leeuw 2005; Groves 2004). The impact of interviewer presence can be seen in a number of studies comparing interviewer-administered and self-administered modes in studies of sensitive health behaviors where self-administered modes have resulted in higher reported rates of these behaviors (Cooley et al. 2000; Aquilino 1992; Turner et al. 1998; Tourangeau, Steiger, and Wilson 2002; Tourangeau and Smith 1996). As noted above, one implication of such findings has been an increasing trend toward self-administration of sensitive survey questions, even in traditionally interviewer-administered modes.

This study examines whether it is possible to take advantage of the time and cost efficiencies of a web mode of survey administration while minimizing coverage and nonresponse error by offering in-person interviews for participants who lack Internet access or are reluctant to use the web mode. Further, we examine whether responses to questions concerning sexual activity and drug use differ by mode. Participants in the current study were randomly assigned to two conditions of mixed-mode approaches to data collection. One condition started with a web survey and followed with an in-person interview for noncompleters after a specified period of

time, and the other condition began with an in-person interview and followed with a web survey for noncompleters. In this study we compare the two conditions with respect to cost, time, and error. In analyzing nonresponse and coverage error, overall completion rates are compared for the two conditions, as well as completion rates and rates of compliance with mode assignment for subgroups of the sample. With respect to measurement error, comparisons are made by mixed-mode condition, examining rates of incomplete data and means or prevalences of responses to survey measures of sexual risk behavior and substance use.

Methods

Design & Sample

The data used for this study come from Raising Healthy Children (RHC), a longitudinal study of the development of substance use, other problem behavior, and positive behavior. Nested within the project is an experimental test of a social development approach to prevention, with half the sample assigned to the intervention condition (for more details of the intervention see Catalano et al. 2003; Brown et al. 2005; Haggerty et al. 2006). The study population originally consisted of first- and second-grade students in ten suburban public elementary schools in a Pacific Northwest school district. To be included in the RHC sample, students had to remain in their school throughout the entire first year of their participation in the study and have a parent who spoke English, Spanish, Korean, or Vietnamese. In Year 1, 938 parents of 1,239 eligible students provided written consent to participate in the study. In Year 2, the sample was augmented with an additional 102 students from a second pool of 131 eligible students who newly entered one of the ten schools during second grade, thus yielding a total sample of 1,040 students. These students were interviewed each spring through the twelfth grade. When participants turned eighteen, they were asked to provide written consent to continue participation in the study. In 2004, RHC surveyed the older cohort of participants ($n = 501$) during the fall after high school graduation. This older cohort constitutes the potential sample for the present investigation.

The retention rate for the oldest cohort has been high through high school, with survey completion rates ranging from over 95% for the elementary school surveys to 89% in twelfth grade. The first criterion for inclusion in the current study was that participants were still active in the RHC project, removing forty-two participants who withdrew from the project prior to the fall post-high school survey. The second criterion was that participants had to live within Washington State at the time of the fall survey, because all out-of-state participants were first asked to complete the survey via Web and only those participants who lived in Washington were randomly assigned to one of the two mixed-mode conditions. This second inclusion criteria removed another seventy-three participants from the sample, leaving an analysis sample of 386.

The average age of youth participants in September 2004 was 18.56 years (range 17.56 – 19.79). Fifty-five percent were in the intervention condition of the experimental test of a social development prevention program, and 58% were male. The ethnic/racial composition of the sample was 82% White, 4% Hispanic, 3% African American, 3% Native American, and 8% Asian or Pacific Islander. At the beginning of the RHC study, 26% were living in low-income families, indicated by whether the child was enrolled in the free or reduced-price school lunch program in the first two years of the project.

Procedures

Study procedures were approved by a University of Washington Human Subjects Review Committee. Since respondents were in the tenth grade, annual spring surveys have been administered one-on-one by interviewers using laptops with self-administered sections for

sensitive items. Throughout the study, small gifts (e.g., a clock radio) or monetary incentives were given to respondents who completed their surveys. For completing the survey administered during the fall of 2004, youth received a \$20 cash incentive.

The fall survey included 274 items, although, depending on skip patterns, a respondent might be asked to answer as few as 158 items. Questions asked for information on sex risk behavior, substance use, and respondents' social environment (including participants' work and school experiences and their relationships with peers, family, and intimate partners). Most questions used a time frame of the prior thirty days. The median amount of time to complete a survey was thirty-two minutes for the in-person mode and thirty-one minutes for the web mode.

Participants residing in Washington State were stratified by intervention condition, gender, and whether they were attending college, and were then assigned randomly to one of two mixed-mode conditions: in-person interview first with web survey follow-up, or web survey first with in-person follow-up. For both conditions, there was a rolling release date, with participants released into the field at one of four dates: 10/1/04, 10/16/04, 10/21/04, and 10/30/04. This process spread out the workload for interviewing staff. Study participants who indicated during the prior summer survey administration that they would be attending college in the fall and who volunteered their school's start date were assigned to a release group so that their interview would take place no less than 30 days after they began college. The original goal was to have participants complete surveys within 62 days of their release date, although the field period lasted until the end of January in order to achieve the overall target completion rate of 85%.

Both conditions had a standardized contact strategy. Advance letters were mailed to participants in both conditions on their release date. The web-first condition received an email reminder six and fourteen days after their release into the field, with telephone follow-up beginning twenty-five days after the release date. At the time of the telephone follow-up, participants in the web-first condition were informed that they had the option of completing their interview in the in-person mode, although they were encouraged to complete their survey on the Internet.

Participants in the in-person-first condition were assigned to interviewers at the release date and interviewers contacted them by phone, email, and mail in order to schedule interviews. The frequency, type, and number of these contacts were determined on a case-by-case basis. Late in the field period (December), participants in the in-person-first condition were offered the option of completing their surveys by the web mode.

In the web mode, study participants were asked to use a point-and-click procedure to select their responses to all questions. For most items, questions were asked in a table format. Of the 274 items in the survey, 173 were not required (web respondents could move to the next item by simply clicking on the "next" button) while 103 were required (respondents had to answer the question or provide an active response "don't know," "not applicable," or "refused"). In general, the items that required an "active" nonresponse were items that activated skip rules for subsequent questions.

In the interviewer-administered mode, nonsensitive questions were asked aloud by interviewers who recorded answers on a laptop computer. For many interviewer-administered questions in the in-person interview mode, interviewers used show cards with the response categories numbered or lettered so that respondents only had to say a letter or number reflecting their response, not the actual response choice. To further minimize social desirability effects, sensitive questions about young peoples' own behavior were collected in a computer-assisted self-administered (CASI) format in which the laptop was turned over to the respondent to read the questions and enter his/her responses. All items in the self-administered portion required

that a respondent either answer the question or push a function button, followed by choosing “don't know,” “refuse,” or “not applicable.”

In order to prevent partially completed surveys, the \$20 incentive was contingent on a participant completing the survey. Young adults who completed the survey in person received the incentive from the interviewer if they went through the entire survey, either answering or choosing not to answer each question. If a participant was unable to finish the in-person interview in one sitting, study staff attempted to reschedule an appointment so that the participant could complete the interview and receive the incentive. For those who initiated a web survey and broke off the survey in the middle, participants were instructed to log on again and continue with the interview where they left off. Follow-up telephone calls were made to these participants to encourage them to log back in and complete their surveys. Participants who completed the web survey received an incentive by mail if they went through the entire survey and clicked on the “submit” button at the end.

Measures

The monetized cost per interview completed was calculated in 2004 dollars, based on the aggregate costs of locating, interviewer time, mileage, cell phones, postage, hosting fees, programming time, supervision time, and incentives. The cost estimates for the two conditions do not include photocopies, supplies, or software costs (e.g., the CAPI and web programming software) which were comparable between conditions and modes. Costs of interviewer time were based on the number of telephone, in-person, and email contacts made per completed interview. The average time per each of these types of contact attempts was based on a record of minutes per each type of attempt recorded, and these average times were multiplied by average interviewer compensation rates in 2004.

Time until completion—was calculated for each participant who completed a survey or interview based on the number of days after the participant's release date that the interview or survey was completed.

Nonresponse—was based on whether a participant failed to begin a survey. A survey was considered initiated if a respondent signed the consent form, began the survey, and responded to at least one survey item by the end of the field period.

Compliance—with mode assignment was based on whether participants initiated a survey in the mode they were first offered.

Two types of incomplete data for those who initiated surveys were examined. First, partial interviews were counted when a participant began a survey but broke off at some point, answering no items after that point in the survey. For the web mode, these partial surveys involved participants closing out the survey and never logging back in to complete all the items. For in-person interviews, this involved not completing the interview in one sitting and failing to make a rescheduled appointment to complete the rest of the interview. Second, among those who completed a survey, nonresponse to individual items was calculated based on the number of items that were skipped or to which respondents answered “refused,” “don't know,” or “not applicable.”

In order to examine response bias for items concerning sensitive behaviors, prevalence or means on survey measures were examined across condition and completed mode, first for items and measures that were in the self-administered portion of the surveys completed in person, and second, for items or measures that were in the interviewer-administered portion of the in-person interview.

Among items and measures in the self-administered section of the in-person survey were fourteen individual items concerning substance use and sexual risk behavior that were dichotomized to produce rates of reported behavior. Seven measures were based on multiple component scales. Specifically, three scales measured sex-related alcohol expectancies (i.e., beliefs about alcohol enhancing sex, increasing risky sex, and decreasing inhibition--Dermen and Cooper 1994), and three scales measured sex-related drug expectancies (i.e., beliefs about drugs [like marijuana, cocaine, or ecstasy] enhancing sex, increasing risky sex, and decreasing inhibition--adapted from Dermen and Cooper (1994)). The other composite measure was the Rutgers Alcohol Problem Index (RAPI; White & Labouvie (1989), calculated as the sum of responses to seventeen alcohol-caused problems experienced in the prior thirty days.

Among measures asked in the interviewer-administered sections of the in-person survey were two individual items (whether a family member or romantic partner had a drinking or drug problem) and six scales. Four of these scales concerned romantic partner's and peers' substance use, peer sexual behavior, and peers' encouragement of substance use and risky sexual behavior. The other two scales measured general expectancies around alcohol and marijuana use (Brown, Christiansen, and Goldman 1987; Brown, Greenbaum, and Maller 1999).

Analysis Strategy

Cost per interview was based on aggregate cost for participants in each condition and does not lend itself to a statistical test of a null hypothesis. Comparison of nonresponse rates and time to completion were examined with assigned condition treated as the independent variable. Comparison of compliance with assigned mode was conducted separately for the two conditions, with predictors of compliance treated as independent variables in separate models.

In order to explore mixed-mode effects on measurement error, we relied primarily on intent-to-treat analyses comparing all participants by randomly assigned condition. First, we compared rates of partial interviews and nonresponse to individual items across assigned modes. Second, we examined prevalences and means on survey measures. These analyses used logistic regression models for the dichotomous measures and ANCOVA models for mean levels on composite measures, in both cases controlling for variables we have found in prior analyses to be related to rates of sexual risk behavior and substance use in this developmental time period. Control variables were gender (male vs. female), residential status (left home vs. living with parents), and, when available, prior report for the specific measure from the spring 2004 survey. Alternate models were run with additional control variables, such as college status and experimental condition, but these additional control variables explained little additional variance in the dependent variables examined and their presence did not change the direction or statistical significance of relationships with assigned mode.

This study could be seen as an example of a randomized design with noncompliance in both experimental, mixed-mode conditions, wherein some participants assigned to each mixed-mode condition were able to choose the intended experience of the other condition. Secondary analyses were run using "as-treated analyses" that compared responses by completed mode, with the caveat that such models may result in biased results due to pooling the responses of participants who complied with their assigned condition with those assigned to the other condition who did not comply (Freedman 2006). The results of these models should be interpreted as exploratory analyses of possible mode effects. Instrumental variable (Freedman 2006; Angrist, Imbens, and Rubin 1996) models that use information on compliance rates in the two conditions in order to estimate the effect of the "treatment" on those who were actually treated, were also considered. In the current study, results of intent-to-treat and instrumental variable models were consistent in terms of demonstrating similar significance levels for estimated treatment effects. However, the case may be made that these data do not meet key assumptions of the instrumental variable model because, as reported below, assigned mode

was related to how soon in the field period participants completed their interviews. Thus, it may be that assigned mode affected responses through the timing of the interviews, suggesting a violation of the *exclusion restriction* that the effect of assignment on outcomes operates entirely through the “treatment” (Angrist, Imbens, and Rubin 1996). We therefore chose to present the results of the intent-to-treat analyses as the primary focus of this paper because they are consistent with the randomized nature of the overall study design.

Results

Cost

The cost per interview completed was \$72 for those assigned to the web-first condition and \$114 for those assigned to in-person-first condition. Most of the cost difference was due to more respondents in the in-person-first condition receiving in-person visits from an interviewer compared to participants in the web-first condition (88% vs. 24%, $X^2 = 144.15$, $p < .01$). Also, participants in the in-person-first condition required a mean of 7.71 telephone contacts ($SD = 8.15$) compared to a mean of 4.68 ($SD = 7.30$) for those in the web-first condition ($t = 3.71$, $p < .01$). In-person-first participants were less likely to be contacted by email (27% vs. 39%, $X^2 = 6.70$, $p < .01$), although there were minimal costs associated with email contacts.

Time

Figure 1 shows the cumulative percentage of participants in each mixed-mode condition who completed their surveys by days after release date, while Figure 2 shows the percentages of participants in each condition who completed interviews within 20-day intervals after release date. As shown in the two figures, more interviews were completed earlier in the field period in the in-person-first condition. The average time until completion among participants assigned to the in-person-first condition was 22.94 days ($SD = 23.54$) compared to 34.30 days ($SD = 26.53$) for those assigned to the web-first condition ($t = 4.27$, $p < .01$). By day 25 of the field period (which is the time at which telephone follow-up calls began for those assigned to the web-first condition), 42% of the web-first interviews were completed compared to 71% of the in-person-first interviews. By the two-month target date for survey completions, the gap in completions had narrowed but still favored the in-person-first condition, with 79% of web-first and 87% of in-person-first participants completing surveys ($X^2 = 3.92$, $p < .05$).

Nonresponse

By the end of the field period, nonresponse rates were similar across assigned conditions, with less than 8% nonresponse in both conditions (see Table 1). The difference in nonresponse rates was not statistically significant ($X^2 = 0.55$, $p > .05$). Among females, nonresponse rates were slightly higher for those assigned to in-person-first condition than those assigned to web-first, although this difference was not statistically significant (5% vs. 10%, $X^2 = 1.78$, $p > .05$). Among males there was a statistically significant difference, with nonresponse more common in the web-first condition (9% vs. 2%, $X^2 = 5.49$, $p < .05$). Differences in nonresponse rates by assigned condition were nonsignificant for participants in both the initial intervention and control groups for the RHC study.

Compliance

There were some notable differences with respect to mode compliance that reflect differential coverage or nonresponse bias of the web mode. Among those assigned to the web-first condition, females were more likely than males to complete by Web (80% vs. 60%, $X^2 = 8.57$, $p < .01$), as were individuals in the control condition of the RHC study versus those in the intervention condition (76% vs. 62%, $X^2 = 4.27$, $p < .05$) and those enrolled in school versus those not in school (82% vs. 61%, $X^2 = 9.39$, $p < .01$). Among those assigned to the in-person-

first condition, females were again more likely than males to complete by Web (17% vs. 7 %, $X^2 = 4.15$, $p < .05$). A small number of participants in both the in-person-first and web-first conditions completed their interviews by telephone.

Incomplete Data

In general, incomplete data were minimal. There were only three partial interviews among the entire analysis sample, all of them involving individuals that began filling out the survey on the Web, broke off, and failed to log back in to complete the survey. Two of these individuals had originally been assigned to the in-person-first condition, while the other was in the web-first condition. Among those completing the interview, there were no statistically significant differences in the number of refused or skipped questions by mixed-mode condition. There was an average of 1.89 (SD = 3.04) items with missing data for those assigned to the web-first condition, compared to 2.30 (SD = 3.55) items for those assigned to in-person-first condition ($t = 1.18$, $p > .05$). Secondary analyses comparing surveys completed via Web with those completed in person did, however, suggest some mode impacts on incomplete data. Individuals who completed the interview by Web did not answer an average of 1.19 items (SD. = 1.95), while those who completed the interview in person did not answer an average of 2.77 (SD = 3.91) items ($t = 4.54$, $p < .01$). The source of this difference was more frequent use of the “don't know” option by participants completing the survey in person (mean = 2.07 for in-person vs. mean = .35 for web, $t = 5.42$, $p < .01$).

Response Bias

Table 2 presents the results of comparing assigned (in-person-first and web-first) mode across measures of substance use, sexual behavior, sex-related alcohol and drug expectancies, and alcohol problems. These measures were self-administered in both web and in-person modes. Comparing assigned mixed-mode conditions in the first two columns of Table 2, there was one significant difference in the 29 measures of drug use, sexual behavior, and expectancies after adjusting for control variables. Young adults in the in-person-first condition were more likely to report using marijuana in the past month compared to those in the web-first condition.

Table 3 presents the results of comparing responses to questions about the drug use and sexual behavior of family members, romantic partners, and friends; peer encouragement; and positive effects of alcohol and marijuana. These questions were self-administered on the web survey but administered by an interviewer on the in-person interview; thus, these results represent a critical test of where administration mode effects might create differences between the two conditions. No significant differences were found in the assigned mode comparison, although the difference in report of a family member having a drinking problem approached significance, with web survey completers more likely to respond affirmatively to this item. Secondary analyses that compared in-person completers with web completers indicated a significant difference on this item by completed mode (15% vs. 26%, $p < .05$).

Discussion

The current study describes a randomized controlled study of mixed modes of data collection. The underlying goal was to assess whether it was possible to minimize costs of data collection and shorten field periods by using a web survey mode in conjunction with in-person surveys, while at the same time minimizing survey error. In general, results of the study support the use of a mixed-mode approach.

The cost savings of using a web-based mode of data collection with in-person backup were clear. Due to less data collection time spent on travel to in-person interviews and fewer telephone calls to arrange in-person appointments, the cost per completed interview was 46%

less for the condition in which web interviews were offered first with in-person backup. Such a savings supports the point that a web-first, mixed-mode data collection strategy can result in substantial cost savings to survey researchers (Groves 2004).

Effects on the length of the field period, however, were not what we expected. Participants assigned to the web-first condition were, in general, not quick to jump at the chance to complete their interviews over the Internet and needed to be encouraged to do so. In hindsight, the slower time to completion for those in the web-first condition was likely the result of not assigning participants in this condition to interviewers until late in the field period (i.e., day 25). Follow-up calls and emails, tailored to each individual case, earlier in the field period may have decreased this lag in completion times. Also, as participants become more familiar with completing surveys over the Internet, this hesitancy to try this new survey mode might diminish. Nonetheless, shortening the length of field periods remains a challenge; offering the web survey mode with in-person back-up did not shorten the field period for this study.

Although we failed to meet our completion rate goals within 62 days of the beginning of the field period, overall completion rates were high for both mixed-mode conditions, suggesting that coverage and nonresponse errors were minimized through using the mixed-mode strategy. Coverage and nonresponse bias with using only a web survey were strongly suggested by the fact that, among those assigned to the web-first condition, male participants in the intervention condition of the RHC study and participants not attending school were less likely than their counterparts to complete surveys via the Web. For nonstudents, this difference likely reflects less access to the Internet. Some evidence of nonresponse and coverage error for the in-person mode also exists, as indicated by two findings: 1) adequate completion rates for the in-person-first mode were only achieved when the web survey was offered to difficult-to-interview participants, and 2) females assigned to the in-person-first condition were more likely than males not to comply with their assigned mode and complete the survey via the Web. In sum, our findings with respect to compliance with assigned mode in a mixed-mode format suggest that completed mode may be associated with important background variables, such as gender, experimental condition, and college status.

Evidence from the current study suggests that incomplete data can be kept to a minimum, regardless of the mixed-mode strategy used for data collection. Only in a few cases were participants who completed surveys in person slightly more likely to use the “don't know” response to questions. We believe this may be related to question formatting. In the web mode, the “don't know” option did not appear on the screen as one of the response options. Only after a participant tried to skip out of a question by clicking on the “next” button without indicating an item response was the “don't know” option offered. In contrast, the show cards with response options given to participants completing surveys in-person included a “don't know” option as part of the response set for each question.

The current study also investigated response equivalence by assigned mode. Overall, there were minimal statistically significant differences in responses to questions about drug use and sexual behavior between web surveys and in-person interviews. Of the twenty-nine assigned condition comparisons, only one was statistically significant at the $p < .05$ level, which could be expected by chance alone. We do note that differences found by completed interview in reports of family drug use suggest a plausible administration mode effect on this item, with respondents who completed the interview in-person perhaps being less likely to say aloud to the interviewer that they had a family member with a drinking problem.

Although we found few significant differences between mixed-mode conditions, it is important to identify caveats of the current study. First, this study is a comparison of mixed-mode strategies in the context of a larger longitudinal study. Although the in-state sample was

randomly assigned to a mixed-mode condition, response rate goals for the larger study meant study staff exercised a great deal of persistence and flexibility to achieve survey completion. Second, this is not a first-time sample of young adults, but an established sample of respondents who were accustomed to being interviewed on an annual basis by trained interviewers. Thus, results are likely to be less generalizable to other types of samples. However, because we have previous reports from these young adults, we are able to control for reports of prior behaviors, strengthening the conclusions of this study regarding equivalence of responses to sensitive questions. Third, the sample size, although moderate, may have limited our power to detect significant differences. Some of the statistically nonsignificant differences in reported prevalences by condition are of a magnitude that may be substantively important. Finally, in order to minimize nonresponse, participants in the web-first mode were eventually allowed to complete their interviews in person, those in the in-person-first mode were eventually allowed to complete their interviews by Web, and in both conditions, a small number of participants completed their interviews by phone. Thus, this study is not a direct test of interview mode effects on response bias, but rather a randomized study of two different mixed-mode conditions. Secondary analyses that compared participants' responses by completed modes or "as-treated" are not consistent with the randomized design of the study and may be subject to biased results due to plausible violations of model assumptions (Freedman 2006). Results of such secondary analyses should be interpreted as exploratory.

Evidence continues to accumulate on the quality and effectiveness of web-based surveys. For researchers designing longitudinal data collection strategies to maximize efficiency, the utility of being able to offer multiple high-quality data collection modes that include web surveys is appealing to both maximize participation rates and minimize costs. Experiments on the impact of mixed-mode surveys on completion and responses to sensitive questions provide important contributions to knowledge about efficiency and quality of mixed-mode approaches.

Acknowledgements

This research was supported by research grant # R01 DA08093-13 from the National Institute on Drug Abuse. A previous version of this paper was presented at the Society for Prevention Research Annual Meeting, Washington, DC, May 2005.

References

- Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 1996;91:444–455.
- Aquilino WS. Telephone versus face-to-face interviewing for household drug use surveys. *International Journal of the Addictions* 1992;27:71–91. [PubMed: 1537641]
- Brown EC, Catalano RF, Fleming CB, Haggerty KP, Abbott RD. Adolescent substance use outcomes in the Raising Healthy Children Project: A two-part latent growth curve analysis. *Journal of Consulting and Clinical Psychology* 2005;73:699–710. [PubMed: 16173857]
- Brown, EC.; Greenbaum, PE.; Maller, SJ. Using item response theory to improve measurement of alcohol expectancies; Paper read at 22nd Annual Scientific Meeting of the Research Society on Alcoholism; Santa Barbara, CA. 1999.
- Brown SA, Christiansen BA, Goldman MS. The Alcohol Expectancy Questionnaire: An instrument for the assessment of adolescent and adult alcohol expectancies. *Journal of Studies on Alcohol* 1987;48:483–491. [PubMed: 3669677]
- Catalano RF, Mazza JJ, Harachi TW, Abbott RD, Haggerty KP, Fleming CB. Raising healthy children through enhancing social development in elementary school: Results after 1.5 years. *Journal of School Psychology* 2003;41:143–164.
- Clayton, RL.; Werking, GS. Business surveys of the future: The World Wide Web as a data collection methodology. In *Computer assisted survey information collection*. In: Couper, M.; Baker, RP.;

- Bethlehem, J.; Clark, CZF.; Martin, J.; Nicholls, WL., 11; O'Reilly, JM., editors. Wiley; New York: 1998. p. 543-562.
- Cooley PC, Miller HG, Gribble JN, Turner CF. Automating telephone surveys: using T-ACASI to obtain data on sensitive topics. *Computers in Human Behavior* 2000;16:1-11.
- Couper MP. Web surveys: A review of issues and approaches. *Public Opinion Quarterly* 2000;64:464-494. [PubMed: 11171027]
- de Leeuw ED. To mix or not to mix data collection modes in surveys. *Journal of Official Statistics* 2005;21:233-255.
- Dermen KH, Cooper ML. Sex-related alcohol expectancies among adolescents: II. Prediction of drinking in social and sexual situations. *Psychology of Addictive Behaviors* 1994;8:161-168.
- Dillman, DA. Mail and internet surveys: The Tailored Design Method. Wiley; New York: 2000.
- Dillman DA, Christian LM. Survey mode as a source of instability in responses across surveys. *Field Methods* 2005;17:30-52.
- Freedman DA. Statistical models for causation: What inferential leverage do they provide? *Evaluation Review* 2006;30:691-713. [PubMed: 17093105]
- Groves, RM. Survey errors and survey costs. Wiley-Interscience; Hoboken, New Jersey: 2004.
- Groves, RM.; Couper, MP. Nonresponse in household interview surveys. Wiley; New York: 1998.
- Haggerty KP, Fleming CB, Catalano RF, Harachi TW, Abbott RD. Raising Healthy Children: Examining the impact of promoting healthy driving behavior within a social development intervention. *Prevention Science* 2006;7:257-267. [PubMed: 16752099]
- Kraus L, Augustin R. Measuring alcohol consumption and alcohol-related problems: Comparison of responses from self-administered questionnaires and telephone interviews. *Addiction* 2001;96:459-471. [PubMed: 11255585]
- Krosnick JA. Survey research. *Annual Review of Psychology* 1999;50:537-67.
- Kwak N, Radler B. A comparison between mail and web surveys: Response pattern, respondent profile, and data quality. *Journal of Official Statistics* 2002;18:257-273.
- Kypri K, Gallagher SJ, Cashell-Smith ML. An Internet-based survey method for college student drinking research. *Drug and Alcohol Dependence* 2004;76:45-53. [PubMed: 15380288]
- Leece P. Correction and republication: Internet versus mailed questionnaires: A controlled [correction of "randomized"] comparison (2). *Journal of Medical Internet Research* 2004;6:e38.
- McCabe SE, Boyd CJ, Couper MP, Crawford S, D'Arcy H. Mode effects for collecting alcohol and other drug use data: Web and U.S. mail. *Journal of Studies on Alcohol* 2002;63:755-761. [PubMed: 12529076]
- Parks KA, Pardi AM, Bradizza CM. Collecting data on alcohol use and alcohol-related victimization: a comparison of telephone and Web-based survey methods. *Journal of Studies on Alcohol* 2006;67:318-323. [PubMed: 16562415]
- Schleyer TKL, Forrest JL. Methods for the design and administration of web-based surveys. *Journal of the American Medical Informatics Association* 2000;7:416-425. [PubMed: 10887169]
- Shannon DM, Bradshaw CC. A comparison of response rate, response time, and costs of mail and electronic surveys. *Journal of Experimental Education* 2002;70:179-92.
- Singer E. Nonresponse bias in household surveys. *Public Opinion Quarterly* 2006;70:637-645.
- Steeh CG. Trends in nonresponse rates, 1952-1979. *Public Opinion Quarterly* 1981;45:40-57.
- Tourangeau R, Smith TW. Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly* 1996;60:275-304.
- Tourangeau R, Steiger DM, Wilson D. Self-administered questions by telephone: Evaluating interactive voice responses. *Public Opinion Quarterly* 2002;66:265-278.
- Turner CF, Ku L, Rogers SM, Lindberg LD, Pleck JH, Sonenstein FL. Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science* 1998;280:867-873. [PubMed: 9572724]
- Weisberg, HF. The total survey error approach: A guide to the new science of survey research. University of Chicago Press; Chicago: 2005.
- White HR, Labouvie EW. Towards the assessment of adolescent problem drinking. *Journal of Studies on Alcohol* 1989;50:30-37. [PubMed: 2927120]

Figure 1.
Cumulative percent of interviews completed by days after release date for in-person-first and web-first conditions.

Figure 2.
Histogram of when interviews completed for in-person-first and web-first conditions.

Table 1

Assignment and Completion Rates

	Assigned in-person first	Assigned web first	Totals
Completed in person	156 (83%)	45 (23%)	201 (52%)
Completed via Web	21 (11%)	135 (67%)	156 (40%)
Completed by phone	2 (1%)	3 (2%)	5 (1%)
Non-completes	10 (5%)	14 (7%)	24 (6%)
Totals	189 (100%)	197 (100%)	386 (100%)

Table 2

Reports of Respondents' Own Drug and Sexual Behavior and Expectancies by Mixed-Mode Assignment

	In-person first n = 178 %	Web first n = 181 %	P
Single items			
<i>In the past 30 days...</i>			
Drank alcohol	59.3	55.0	.692
Drank >5 drinks on one occasion	35.0	41.7	.837
Used marijuana	37.3	27.6	.018
Used cocaine	5.1	6.6	.472
Used other opiates	13.0	11.0	.371
Drove car after drinking/drug use	16.4	11.0	.172
Rode in a car w/someone drinking or using drugs	23.2	19.3	
Had sex	55.1	55.9	.594
Drank alcohol prior to sex	33.3	46.2	.071
Used drugs prior to sex	17.8	25.8	.154
Had >1 sex partner	5.1	7.3	.401
Used condoms	60.5	61.9	.402
Used other birth control	74.4	75.0	.744
<i>Ever had a sexually transmitted disease (STD)</i>	5.1	6.8	.568
Multiple-item measures	M (SD)	M (SD)	
<i>Expectancies for alcohol...</i>			
Enhancing sex	3.01 (0.77)	3.06 (0.96)	.901
Increasing risky sex	3.73 (0.87)	3.84 (0.94)	.440
Decreasing inhibition	3.73 (0.86)	3.80 (0.78)	.413
<i>Expectancies for drugs...</i>			
Enhancing sex	2.98 (0.86)	3.04 (0.89)	.495
Increasing risky sex	3.53 (0.96)	3.62 (0.95)	.350
Decreasing inhibition	3.53 (0.92)	3.52 (0.88)	.847
<i>Rutgers Alcohol Problems Index</i>	1.67 (3.80)	1.34 (2.68)	.101

Notes: In the in-person interview mode, these questions were self-administered on laptop computers. * $p < .05$. Significance levels of comparisons by mixed-mode condition assessed by logistic or linear regression models that adjusted for gender, living at home with parents or not, and previous report of behavior when available on the spring 2004 survey. Means and standard deviations are from unadjusted estimates.

Table 3

Reports of Others' Drug and Sexual Behaviors, Encouragement, and Drug Expectancies by Mixed-Mode Assignment

	In-person first n = 178 %	Web first n = 181 %	P
Single items			
Family member has drinking problem	16.0	23.9	.072
Boy/girlfriend/spouse has drinking/drug problem	8.5	5.9	.722
Multiple-item measures			
	M (SD)	M (SD)	
Boy/girlfriend/spouse's drug use	1.67 (0.69)	1.64 (0.68)	.863
Peer drug use	2.11 (0.90)	2.12 (0.89)	.919
Peer sexual behavior	1.88 (0.90)	1.82 (0.77)	.396
Peer encouragement of drug use and sex	1.38 (0.54)	1.49 (0.60)	.193
Alcohol expectancies	2.68 (0.55)	2.65 (0.68)	.984
Marijuana expectancies	2.72 (0.59)	2.66 (0.65)	.444

Notes: * $p < .05$. Significance levels of comparisons by mixed-mode condition assessed by logistic or linear regression models that adjusted for gender, living at home with parents or not, and previous report of behavior when available on the spring 2004 survey. Means and standard deviations are from unadjusted estimates.