Available online at www.sciencedirect.com
ScienceDirect
Computers in Human Behavior 24 (2008) 903–916
Computers in Human Behavior
www.elsevier.com/locate/comphumbeh

The impact of computer versus paper–pencil survey, and individual versus group administration, on self-reports of sensitive behaviors

Scott C. Bates *, Jared M. Cox

Department of Psychology, Utah State University, 2810 Old Main Hill, Logan, UT 84322-2810, United States

Available online 25 May 2007

Abstract

The purpose of this study was to clarify the strengths and weaknesses of both computer based and paper–pencil methods of sensitive data collection in various administrative settings. A total of 180 students signed-up for optional course-credit and were randomly assigned to complete a questionnaire using either a paper–pencil, or a computer based method, and in one of the following administration settings: group, alone in a designated office, or being emailed/mailed the information and filling it out when/where they pleased. Results show that perceptions of anonymity and confidentiality were strongly correlated with self-reported accuracy. However, although perceptions of anonymity and confidentiality differed by condition, this had no statistically significant impact on responses. Though perceptions of anonymity and confidentiality differ between methods, neither method appears to be superior in the context of overall response bias.

Keywords: Computerized questionnaires; Sensitive information disclosure; Paper and pencil questionnaires; Anonymity

1. Introduction

The *No Child Left Behind Act of 2001* specified parameters under which states would be mandated to monitor alcohol, tobacco, and other drug (ATOD) use among adolescents.

* Corresponding author. Tel.: +1 435 797 2975; fax: +1 435 797 1448.
E-mail address: scottcbates@cc.usu.edu (S.C. Bates).

0747-5632/$ - see front matter
doi:10.1016/j.chb.2007.02.021

While adolescent ATOD use has been systematically monitored at the national level since the 1970s (Johnston, O'Malley, & Bachman, 2003), many states and local education agencies have not been systematically monitoring adolescent ATOD use. The most common way to collect ATOD data from adolescents is the ubiquitous school-survey. The use of standard, paper–pencil (PP) methodology to collect ATOD data has been criticized for a variety of reasons.

Adolescent ATOD use surveys include personal and sensitive questions—questions about illegal and socially discouraged behavior. Because such questions are asked, internal validity and pragmatic issues arise. First, two major threats to the validity of findings from studies in which such sensitive questions are asked are critical: (a) demand characteristics (particularly in an adolescent sample) may have a profound impact on the obtained data; and (b) paper-and-pencil surveys are notorious for eliciting exaggerated responses and inconsistent responses (e.g., respondents who report having used alcohol in the past month, but never have in their lifetimes). Second, three important pragmatic issues also arise: (a) standard paper and pencil (PP) survey methodologies are inflexible when it comes to question-selection, as a printed document cannot be customized at the school building level; (b) collecting data during class is a heavy time burden in an environment where more demands are consistently being put in place; (c) PP methods are expensive in terms of printing, shipping, and processing costs. One solution to the pragmatic issues is on-line data-collection. Less is known, however, about the implications of on-line data-collection on the validity of the data that are collected.

While computers have been used to collect self-report data for some time, the technology and software for collecting data in this manner are still developing. Likewise, research findings regarding their differential impact on data are emerging (see Denscombe, 2006; Walden, 2002).

In most cases, for self-report instruments, higher self-reported rates of sensitive behaviors have been found when computer technology is utilized to collect data. Richman, Kiesler, Weisband, and Drasgow (1999) conducted a meta-analysis of 61 studies published between 1967 and 1997, in which investigators compared the effects of computer-based (CB) surveys, PP surveys, and face-to-face interview data-collection methods on usage rates of sensitive behaviors. While they reported a near-zero overall effect of modality for the comparison between CB and PP survey data-collection methods on general, non-sensitive, questions, they also noted that the quality of the obtained data depended on the administration and design of the CB data-collection instrument. That is, when respondents were alone and could backtrack among the questions, and the anonymity of the participants was ensured, internal validity threats were minor. Richman et al. also found that the impact of data-collection method on internal validity was greater as question sensitivity increased, concluding that more research was needed in the area of CB data-collection when self-reports of sensitive behavior (e.g., drug use) was the dependent variable.

The findings of Richman et al. (1999) have been bolstered by Vereecken and Maes' (2006) recent study on differences between adolescent response styles for CB and PP surveys investigating health and lifestyle behaviors. Vereecken and Maes found that for most questions there were no significant differences. However, for questions of a more sensitive nature, adolescents in the PP condition gave more socially appropriate answers than did those in the CB condition. In a similar study, DiLillo, DeGue, Kras, Di Loreto-Colgan, and Nash (2006) found no significant differences between rates of disclosure of past phys-

ical and sexual abuse using PP, CB, and face-to-face interviews. Their findings did suggest that those in the CB condition were more likely to report distress and mood change than those in the PP and face-to-face interviews. Additionally, the researchers found that participants strongly preferred the CB condition partly due to the confidentiality it could afford them.

A wide variety of investigations in which researchers have used CB self-report surveys of drug use have been conducted. Studies in which investigators examined the effects of data-collection method (CB versus PP) have been conducted with (a) participants undergoing substance-abuse treatment (Sarrazin, Hall, Richards, & Carswell, 2002), (b) adolescent males (Turner et al., 1998), (c) college students (Kypri, Gallagher, & Cashell-Smith, 2004; Mitchell, 1994), (d) adults in an urban setting (Tourangeau & Smith, 1996), and (e) stratified, random, national samples (Wright, Aquilino, & Supple, 1998). Wright et al. (1998), for instance, compared data-collection modality (computer-assisted versus paper-and-pencil) and its impact on findings obtained for a variety of alcohol, tobacco, and other drug use behavior. They collected substance use information using CB and PP methods and compared estimates of self-reported substance use and well-being. Respondents were selected via a national, multi-stage (area) probability sampling procedure; respondents were selected from both urban and suburban areas. Respondents were randomly assigned to an interview mode (PP versus computer assisted (CA)). A total of 3169 interviews were completed; respondents ranged in age from 12 to 34. Interviews were conducted at participants' homes and generally lasted between 45 and 60 min. In the PP condition, respondents were asked to complete a standard drug and alcohol survey while participants in the CA condition were provided a laptop with which to complete the same survey. Wright et al. (1998) reported that across most dependent variables, the findings for adolescents were most sensitive to mode (PP versus CA) as compared to findings for older respondents. For adolescents, higher prevalence rates were consistently reported in the CA mode: twice as many respondents ages 12–18 reported marijuana use in the past 30 days in the CA condition as compared to the PP condition.

While Wright et al. (1998) and Turner et al. (1998) provided valuable insight into possible differences in reported rates of behavior when sensitive questions are involved, there are some fundamental differences between these methodologies used is these two studies and the methodology used to administer standard, school-based, ATOD surveys. Specifically, school-surveys are conducted in classroom settings. Little research has been conducted on the use of computer-based data-collection methodology that would occur in a classroom environment where, given federal directives on adolescent ATOD monitoring, it is likely that data-collection will occur in the future.

The primary purpose of this study was to explore the impact of data-collection methodology on self-reported responses to sensitive questions. Two research questions drove this study: First, would participants' self-reported rates of sensitive behavior (drug use, sexual behavior), or risk-factors associated with sensitive behaviors (perception that drug use is harmful), be impacted (via main-effect or interaction) by data-collection method (CB versus PP) and/or administration setting (group or individual)? Second, would participants' self-reports of accuracy and honesty in their survey responses, and/or their self-reported perceptions of anonymity and confidentiality of their responses, differ as a result of differences in data-collection method and/or administration setting?

2. Method

2.1. Design

A true experiment, utilizing a 2×3 factorial design, was conducted. The independent variables were data-collection method (CB versus PP) and administration setting (group; individual-scheduled; individual-unscheduled). The Group-Scheduled condition replicated the survey administration that occurs in classrooms while both Individual-Scheduled and Individual-Unscheduled conditions served as comparison groups. The Individual-Scheduled and Individual-Unscheduled conditions were important as it is likely that in the future, to comply with federal mandates, data-collection could move outside of the classroom (e.g., in the form of unscheduled, complete-at-home, surveys) and few investigators have studied this form of data-collection in this population.

The primary dependent measures associated with sensitive behaviors included self-report of rates of (a) ATOD use, (b) prevalence of sexual behavior, and (c) rates and types of eating behavior. Additional dependent measures included perceptions of (a) harm associated with ATOD use, (b) confidentiality of participants' responses, (c) anonymity of participants' responses, and (d) self-reported honesty/accuracy in self-reports.

2.2. Participants

A total of 180 participants signed-up for optional course-credit (there were a variety of other options available to them) to participate. Each was randomly assigned to one of the six conditions. The overall response rate, as well as the response rate by method and by setting are included in Table 1. The overall response/participation rate was 76.1% which is generally high for a survey project but understandable given the methodology—participants agreed to participate by signing-up.

Differences by administration methodology and by administration setting were notable. A higher percentage of participants who initially signed-up to participate completed the survey if it were administered using the paper-pencil collection method. In terms of the administration setting, a higher percentage of those who were assigned to the group setting completed a useable survey form. The lowest overall response rate came from the group who was emailed a direct link to the survey in order to complete it on their own time (individual/computer).

There were 137 participants across all conditions. A variety of demographic characteristics were assessed. Within this sample, 86.7% reported being single, 1.5% reported being divorced, and 13.3% reported being married. Most, 91.9%, reporting ethnicity as Caucasian. In terms of sex, 38.2% were male and 61.8% were female. Most participants also reported being fairly early in their collegiate careers: 64.4% reported 3 or fewer semesters

Table 1
Count of participants/response rate by survey administration setting and administration method

	Group	Office	Individual	Overall
Paper/pencil	28/93.3%	21/70.0%	24/80.0%	73/81.1%
Computer	22/73.3%	22/73.3%	20/66.7%	64/71.1%
	50/83.3%	43/71.7%	44/73.3%	137/76.1%

in college. Living arrangements were more varied: 69.4% reported living at home/apartment while 24.6% reporting living in a residence hall and less than 1% reporting living at a fraternity/sorority house. Approximately 95% of the sample belong to the LDS faith which proscribes the use of alcohol, tobacco, illicit drugs and premarital sex. Therefore perceived confidentiality was likely valued for those admitting to indiscretions.

2.3. Refusals/incomplete forms

Some respondents submitted a survey-form, but the form was substantially incomplete. Of the 137 forms that had some answers completed, seven were removed from the data-set due to massive missing-data. These surveys are considered refusals. Refusals were not distributed equally across the survey administration by administration methods groups. None of these refusals came form the individual administration groups (those forms that were mailed, or emailed) and none came from the paper–pencil administration. Among the computer-administration groups, five refusals came from the Individual-unscheduled setting and two from the group setting. Note that paper–pencil refusals that cannot be counted are those that were mailed.

2.4. Instrumentation

The basic instrument from which each of the data-collection conditions were based is the *Youth Risk-Behavior Survey* (Kann, 2001) and the *Communities that Care Youth Survey* (Arthur, Hawkins, Pollard, Catalano, & Baglioni, 2002) and while there is evidence of adequate reliability (Brener et al., 2002) and validity (Hawkins, Arthur, & Catalano, 1997) for these two instruments, additional testing was done to ensure good psychometric characteristics. For the purposes of this experiment, two versions of the complete instrument were created: a PP version and a CB version. Identical questions and response options were used in both versions. Furthermore, the computer version was designed to mimic the features of the paper–pencil version as closely as possible. That is, the computer survey was presented as a series of pages, each page containing multiple questions and there was a visual indication of progress (mimicking the ability of participants responding via PP to see 'how much more he or she has to do'). Additional development activities that took place at that time included scale-development for the following scales: anonymity, confidentiality, and honesty/accuracy.

Both versions included standard demographic questions (age, sex, marital-status, employment-status, ethnic-group) along with questions in the following areas.

2.4.1. ATOD use

Alcohol, tobacco, and other drug use were measured using items that have been established, implemented, and tested in school-based surveys. Examples include: "During your life, how many times have you used or tried one full drink of alcohol (a can of beer, glass of wine, wine cooler, or shot of liquor)?" Both lifetime and past 30-day prevalence items were collected for alcohol, tobacco, marijuana, and 'other illegal drugs.'

2.4.2. Sexual behavior

Prevalence questions that inquire about sexual behavior were included. Sample questions include: "Have you ever had sexual intercourse?" "How old were you when you

had sexual intercourse for the first time?'' ''During your life, with how many partners have you had sexual intercourse?'' and ''During the past three months, with how many partners did you have sexual intercourse?''

2.4.3. Risk factors

The examination of the risk- and protective-factors that are predictive of ATOD involvement has been extensive and ongoing (Hawkins, Catalano, & Miller, 1992). Questions designed to measure respondents' perceptions of the risk of using alcohol, tobacco, and marijuana were included as an initial measure of a risk-factor that has been well established in the literature (Danseco, Kingery, & Coggeshall, 1999). An example item is: ''How harmful do you think it is to use the following substances frequently (daily or almost daily)?''

2.5. Data collection procedures

Participants were asked to volunteer for the study using the standard procedure used to collect data from students enrolled in general psychology. Volunteers were randomly assigned to one of the six conditions (Group-Scheduled, PP; Group-Scheduled, CB; Individual-Scheduled, PP; Individual-Scheduled, CB; Individual-Unscheduled, PP; Individual-Unscheduled, CB).

2.5.1. PP conditions

In group-administration conditions, participants were asked to meet in a classroom on-campus; groups consisted of 35 participants. Administration instructions were then read aloud. Participants were instructed to complete a survey on a variety of risk-behaviors and were informed that their answers would remain anonymous. They were told that once they completed the survey, they were to place their completed survey into the return envelope that the data-collector provided, and to seal the envelope. Once they had submitted their survey they were asked to go to a different (nearby) room for debriefing.

Participants who were assigned to the individual-scheduled condition were instructed to show up at a specific time and asked to complete a survey. Administration instructions were read to each participant orally by the data-collector and he or she was asked to place the survey in a folder that contained other, completed, survey forms.

Participants who were assigned to the individual-unscheduled condition were mailed a survey and asked to return the survey in the envelope provided. Two follow-ups were conducted (once weekly, for two weeks, by telephone) in order to increase the return rate. Standard instructions (same as those used in oral conditions) were included in printed form.

2.5.2. CB conditions

All computers utilized for data-collection used Windows XP operating systems and Microsoft Internet Explorer (version 6) software. In all three computer-administration conditions, the standard administration instructions were presented on the initial screen and also read aloud. Once surveys were completed, participants clicked a 'submit' button.

The group-administration condition was held at an on-campus computer lab. All participants assigned to this group were asked to arrive at the lab at a specified time—the data-collection began after the standard instructions were read aloud (to support the on-screen version).

In the individual-scheduled condition, participants were asked to report to a small office space that contained the computer. Each was greeted by a research assistant, and then asked to complete the survey; participants were left alone to complete the survey.

In the individual-unscheduled condition, participants were emailed a hyperlink address that allowed each to take the survey at a time of his or her choosing. The participants received two email follow-up notices to increase participation.

3. Results

3.1. Factor analyses

Factor analyses using maximum likelihood extraction and varimax rotation were used to identify sets of items that could be used to assess the participant perceptions of (a) anonymity, (b) confidentiality, as well as (c) admissions of honesty/accuracy in the survey.

For the items related to anonymity, the factor analysis converged in four iterations and extracted a single factor that accounted for 76.0% of the variability. The four items, along with corrected item-total correlations, are included in Table 2. Reliability was assessed using Cronbach's alpha to be 0.93.

For the items related to accuracy/honesty of responses, the factor analysis converged in four iterations and extracted a single factor that accounted for 73.3% of the variability. The four items, along with corrected item-total correlations, are included in Table 2. Reliability was assessed using Cronbach's alpha to be 0.92.

Table 2
Scale items and corrected item-total correlations

Scale/item	IT-r
Anonymity	
I do not think it is possible for my identity to be linked to my responses on this survey	0.78
I feel that there is no way for the researchers to link my responses to me	0.84
I am certain that without my name on the survey I am completely anonymous	0.82
I feel confident that there is no way to link my responses to me	0.87
Accuracy	
All of my responses to questions on the survey were honest	0.77
I am confident that I have answered the questions to the best of my ability	0.81
I answered all the questions honestly pertaining to information I felt was personal	0.87
I feel confident in the accuracy of the information I provided	0.78
Confidentiality-1	
I feel confident that my results will be strictly confidential	0.77
I answered all the questions as though my responses would be kept confidential	0.56
I feel this study is completely confidential	0.79
I feel confident that my answers about my sexual behaviors will be strictly confidential	0.86
Confidentiality-2	
I feel confident that my answers about drug use will be strictly confidential	0.75
I answered all the questions knowing that my responses would be completely confidential	0.75
I believe that the laws governing confidentiality will fully protect my identity	0.66
I believe that my responses will not be viewed by friends and relatives	0.53
I completely trust the researchers to keep my answers confidential	0.64
I feel confident that my results will be strictly confidential	0.73

Note. IT-r = item-total correlations (corrected).

For items related to perceived confidentiality, the factor analysis converged in three iterations and extracted two factors. These two factors accounted for 60.4% of the variability in the data. Given a relatively low zero-order correlation between the two extracted factors, both were used in subsequent analyses. The items for each scale, along with corrected item-total correlations, are included in Table 2. Reliability for the first scale was estimated via Cronbach's alpha to be 0.88; for the second scale, it was estimated to be 0.87.

3.2. Scale characteristics

Table 3 shows descriptive statistics of the anonymity, accuracy/honesty and confidentiality scales (note that factor-scores have a mean of 0 and an SD close to 1). Table 4 shows the zero-order correlations among these four scales. Anonymity was correlated with accuracy ($r = 0.33$, $p < 0.001$) and both confidentiality scales $r = 0.62$ ($p < 0.001$) and 0.30 ($p < 0.01$), respectively.

Analyses of variance were conducted on the scales in order to estimate differences in the scales by administration setting (group, individual-scheduled, individual-unscheduled) and administration methods (PP/CB). A significant two-way interaction was detected for the anonymity scale ($F(1,126) = 4.65$, $p <= 0.011$, adjusted R-squared $= 0.07$). Fig. 1 depicts the interaction. Using accuracy as the dependent variable, both main-effects, and the two-way interaction, were non-significant. Using the confidentiality-1 as the dependent variable, both main-effects, and the two-way interaction, were non-significant at the $p = 0.05$ level. Finally, using the confidentiality-2 as the dependent variable, the analysis revealed a statistically significant main-effect for administration method, $F(1,125) = 5.4$, $p = 0.02$. In this case, the CB method was seen as having a higher degree of confidentiality across all three administration settings. The size of the effect, however, was quite weak (eta-squared $= 0.04$).

Table 3
Descriptive data related to the scales

Scale	N	Min	Max	Mean	SD	Skew	Kurtosis
Anonymity	128	−1.13	2.80	0.00	0.97	0.57	−0.49
Accuracy	126	−0.63	3.10	0.00	0.96	1.17	−0.15
Confidentiality-1	126	−2.53	3.06	0.00	0.94	0.83	0.96
Confidentiality-2	126	−1.18	5.61	0.00	0.88	2.31	11.83

Table 4
Correlations among scales

	Anonymity	Accuracy	Confidentiality-1	Confidentiality-2
Anonymity	–			
Accuracy	0.33**	–		
Confidentiality-1	0.62**	0.40**	–	
Confidentiality-2	0.30**	0.21*	0.14	–

* $p < 0.05$.
** $p < 0.01$.

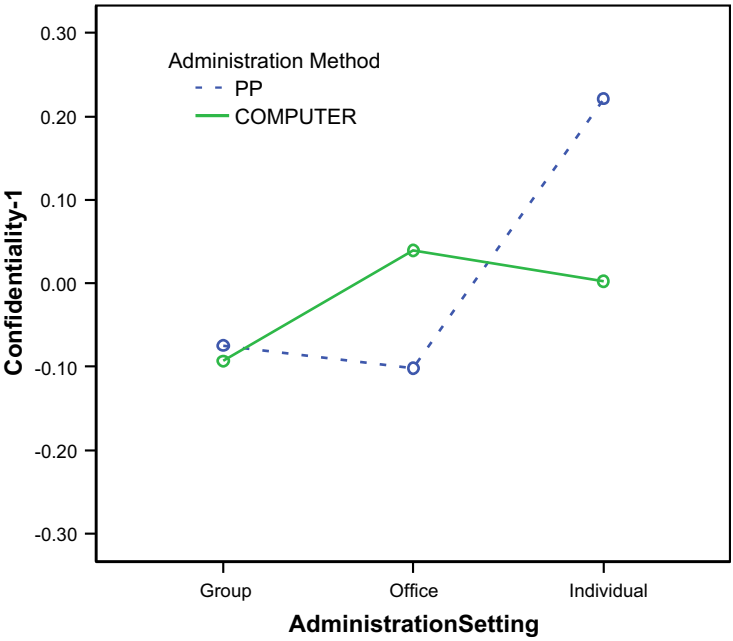

Fig. 1. Estimated confidentiality, by administration setting and method.

3.3. Self-reported behavior prevalence

Five outcome variables were identified as having a high-likelihood for sensitivity to accuracy, confidentiality and accuracy of response. They were: any use of alcohol in the lifetime, alcohol use in the past 30 days, any experience of binge-drinking in the lifetime, having had sexual intercourse at least once, and; having ever given or received oral sex in the lifetime. The Church of Jesus Christ of Latter-day Saints (LDS) explicitly prohibits the use of alcohol as well as sexual intercourse and giving or receiving oral sex prior to marriage (Doctrine & Covenants, 1989). Table 5 shows the prevalence of these behaviors by administration method and setting.

Each of these variables were entered into a logistic regression model in an effort to determine if there were statistically significant differences among the administration setting and administration-method groups. Across all five behaviors, no significant interactions between administration setting or method were detected. However, two main-effects were detected. In the prediction of lifetime prevalence of alcohol use, the model that included administration setting, administration method and their interaction term, a significant difference in setting was revealed: participants were statistically significantly more likely ($p = 0.04$) to indicate having used alcohol in their lifetimes if the data were collected in an individual-unscheduled setting (37.2%) versus an individual-scheduled setting (17.5%). Likewise, in the prediction of lifetime prevalence of having had sexual intercourse, a statistically significant main-effect for administration setting was detected (again, in the context of a complete model). In this case, participants were significantly more likely ($p = 0.05$) to report having had intercourse at least once in their lifetime if their data were

Table 5
Prevalence of behaviors by administration method and administration setting

Variable	Administration method	Administration setting			
		Group	Office	Individual	Subtotal
Alcohol, lifetime	Paper–pencil	28.6	19.0	29.2	26.0
	Computer	35.0	15.8	47.4	32.8
	Subtotal	31.3	17.5[a]	37.2[a]	29.0
Alcohol, 30-day	Paper–pencil	17.9	15.0	21.7	18.3
	Computer	15.8	16.7	15.0	15.8
	Subtotal	17.0	15.8	18.6	17.2
Alcohol, bingeing	Paper–pencil	17.9	4.8	16.7	13.7
	Computer	10.0	11.1	10.0	10.3
	Subtotal	14.6	7.7	13.6	12.2
Sexual intercourse, lifetime	Paper–pencil	28.6	28.6	50.0	35.6
	Computer	30.0	17.6[b]	50.0[b]	33.3
	Subtotal	29.2	23.7	50.0	34.6
Oral sex, lifetime	Paper–pencil	32.1	23.8	50.0	35.6
	Computer	45.0	17.6	42.1	35.7
	Subtotal	37.5	21.1	46.5	35.7

[a,b] Indicate statistically significant differences (based on a multivariate regression model).

collected on an individual-unscheduled basis (either via email or regular mail; 50.0%) as opposed to having the data collected in the individual-scheduled office setting (17.6%).

Next, in an effort to account for differences in perceptions of anonymity and confidentiality, along with these models were re-run with the anonymity, accuracy/honesty and confidentiality scales entered into the model (as a block). Thus, variables were entered as follows: administration method, administration setting, followed by anonymity, accuracy/honesty and confidentiality scales entered in another block.

Using this process to predict lifetime prevalence of alcohol, the significant difference for administration setting remained after anonymity, accuracy/honesty, and confidentiality scales were entered into the predictive model. In addition, accuracy was revealed as a significant, positive, predictor of alcohol use: higher levels of self-reported accuracy were related to higher levels of self-report alcohol use.

The model that best predicted self-reported alcohol in the past 30 days was similarly unchanged—neither administration setting, nor method, nor the interaction between these two variables, was a statistically significant predictor of the outcome variable. However, similar to the prediction model for lifetime alcohol use, self-report accuracy was statistically significant ($p = 0.02$); again, higher levels of self-reported accuracy/honesty were related to higher levels of self-reported alcohol use in the past 30 days.

Taking anonymity, accuracy/honesty and confidentiality into account did change the predictive models for prevalence for bingeing: administration setting again became a significant main-effect. Participants in the individual-unscheduled setting who were mailed the survey (via email or regular mail) were significantly more likely to report alcohol bingeing than if they were administered the survey in the individual-scheduled office setting. Having taken the survey in a group setting was not predictive. In the model for binge-drinking, anonymity, accuracy/honesty and confidentiality scales were not predictive.

The model predicting self-reported sexual intercourse did not change with the addition of anonymity, accuracy/honesty and confidentiality scales into the model and none of these scales significantly predicted self-reported sexual intercourse.

Finally, the model predicted self-reported experiences with oral sex did not change—neither administration setting, nor method, nor anonymity, accuracy/honesty or confidentiality scales were statistically significant predictors of self-reported experiences with oral sex.

4. Discussion

Data were collected from participants who had formally stated that they would participate in the study for class credit. The response rate (76.1%) was not dissimilar from the Kypri et al. (2004) study comparing the response rates for CB versus PP testing methods. That study, using aggressive reminders, had an overall response rate of 82%. However, the current study purposefully targeted a population that would consider the topic of personal alcohol use as a sensitive issue. Additionally, items dealing with what the majority of participants would consider major sexual indiscretions were included in an effort to increase the difficulty that students would have in filling out the survey honestly, or at all.

Most notable are the differences in perceptions of anonymity and confidentiality between administration method for the individual-scheduled and individual-unscheduled settings. The finding that those in the individual-scheduled setting found the CB method of administration to be more anonymous and/or confidential are consistent with those of DiLillo et al. (2006). Their study found that not only did participants disclosing past physical and sexual abuse perceive the office CB condition to be more confidential, but that this condition was also preferable to the office paper pencil condition. The results favoring the individual-unscheduled PP condition over the individual-unscheduled CB condition may suggest a growing concern over internet privacy. The marked increased perception of anonymity/confidentiality reported by participants in the office CB condition suggests that the practice of providing a room and computer may ameliorate privacy concerns.

Although participants' reports of perceived confidentiality and anonymity varied by condition, these perceptions did not translate into significant differences in outcomes. These findings support those of Denscombe (2006) whose study found that for adolescents in group administration settings the method of completion (pencil/paper versus computerized) had little impact on response styles. Further, they are consistent with those of Moore and Ames (2002) who found no significant differences between reports of ATOD use for groups that answered anonymously and those whose surveys lacked anonymity. The current research furthers these lines of investigation by addressing issues of a more sensitive nature, namely alcohol use in an environment that proscribes such use, and sexual improprieties. These results suggest that although certain conditions may be more conducive to increasing perceived anonymity and confidentiality, these factors do not appear to significantly influence responses.

4.1. Implications

There are many advantages to collecting data on computer systems and this study provides some evidence that potential negatives may not be critical. Specifically, though perceptions of confidentiality and anonymity may differ (favoring CB or PP depending on

administration setting) between methods these differences do not appear to impact the nature of the information participants report.

Computer based administration allowed us to directly estimate the number of refusals as data-collection was cumulative. Mailed surveys cannot be assessed for refusals versus other types of response rates. This can provide an important estimate of the generalizability of the results.

The finding that self-reported accuracy was related to perceptions of anonymity and confidentially should not be surprising. The extent to which respondents believe that data are being collected from them in a way that secures their anonymity and confidentiality should predictably impact their reports.

4.2. Limitations

There are two key limitations of this study. The first limitation of this study resides in the fact that the study targeted a population that may not well represent college students in general. Approximately 95% of students at Utah State University belong to the LDS faith which proscribes the use of alcohol, tobacco, illicit drug use, and premarital sexual contact. Use of this population served to effectively increase the uncomfortable nature of the questions creating an atmosphere where anonymity and confidentiality were likely to be highly desirable. However, because these beliefs are not universally shared, their response styles may not be representative of other college campus populations.

The second limitation concerns the scales measuring anonymity and confidentiality. In this case, the rating utilized in this study did not well discriminate perceptions of anonymity and confidentiality between the six conditions. While this may reflect the true lack of significant difference between perceptions, it may also have been due to the scale itself. The five-point Likert-style scales used may not have been sensitive enough to differentiate these perceptions. It may be that a basic threshold level of anonymity and/or confidentiality is requisite to obtaining honest reports about sensitive behaviors. A more sensitive scale may be able to detect the presence of such.

4.3. Future research

Future researchers should focus on the inter-relations among anonymity, confidentiality and self-reported accuracy and honesty. The finding that those in the individual-unscheduled CB condition perceived their anonymity/confidentiality to be less than those in the individual-unscheduled PP condition may reflect general beliefs about internet security. Breeches in computer security are not uncommon and are often presented in the media. This could cause those in the individual-unscheduled CB condition to devalue the use of computers as being confidential and the PP condition may become the more trusted condition.

Additionally, given the data-demands of the k-12 education system as related to alcohol, tobacco and other drug use, and the potentially evolutionary move toward the use of computers to collect these data, it will be important to replicate these results in a younger population. For instance, it is important to understand if having 8th-grade students complete a web-based survey—potentially from home—would impact the accuracy of these results. Likewise, would conducting computer-lab administration, where computer screens

are more readily seen by others who are taking the survey, have as little impact as it did in the present study?

5. Conclusion

As survey methodology continues to evolve, it is important to continue to examine the methodological impacts of these changes on results. As computer and internet access become more ubiquitous, and data-collection becomes more computer-based as a result of the various strengths of that data-collection method, it is important to re-test and revalidate the methods and assess their overall impact on results. With the increase in access to computers and the internet, has come an increase in security issues. These security issues will likely impact participants' perceptions of the anonymity and confidentiality of their data in ways that should be consistently assessed and, where necessary, addressed.

References

Arthur, M. W., Hawkins, J. D., Pollard, J. A., Catalano, R. F., & Baglioni, A. J. (2002). Measuring risk and protective factors for substance use, delinquency, and other adolescent problem behaviors: The Communities That Care Youth Survey. *Evaluation Review, 26*, 575–601.

Brener, N. D., Kann, L., McManus, T. L., Kinchen, S., Sundberg, E. C., & Ross, J. G. (2002). Reliability of the 1999 Youth Risk Behavior Survey Questionnaire. *Journal of Adolescent Health, 31*, 336–342.

Danseco, E. R., Kingery, P. M., & Coggeshall, M. B. (1999). Perceived risk of harm from marijuana use among youth in the USA. *School Psychology International, 20*, 39–56.

Denscombe, M. (2006). Web-based questionnaires and the mode effect: An evaluation based on completion rates and data contents of near-identical questionnaires delivered in different modes. *Social Science Computer Review, 24*, 246–254.

DiLillo, D., DeGue, S., Kras, A., Di Loreto-Colgan, A. R., & Nash, C. (2006). Participant responses to retrospective surveys of child maltreatment: Does mode of assessment matter? *Violence and Victims, 21*, 410–424.

Doctrine and Covenants (1989). Salt Lake City, UT: Church of Jesus Christ of Latter-Day Saints.

Hawkins, J. D., Arthur, M. W., & Catalano, R. F. (1997). *Six state consortium for prevention needs assessment studies: Alcohol and other drugs* (final report for the Center for Substance Abuse Prevention). Seattle, WA: University of Washington, Social Development Research Group.

Hawkins, J. D., Catalano, R. F., & Miller, J. Y. (1992). Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implications for substance abuse prevention. *Psychological Bulletin, 112*, 64–105.

Johnston, L. D., O'Malley, P. M., & Bachman, J. G. (2003). *Monitoring the Future national survey results on drug use, 1975-2002. Vol. I: Secondary school students (NIH Publication No. 03-5375)*. Bethesda, MD: National Institute on Drug Abuse.

Kann, L. (2001). The Youth Risk Behavior Surveillance System: Measuring health-risk behaviors. *American Journal of Health Behavior, 25*, 272–277.

Kypri, K., Gallagher, S. J., & Cashell-Smith, M. L. (2004). An internet-based survey method for college student drinking research. *Drug and Alcohol Dependence, 76*, 45–53.

Mitchell, D. L. (1994). A multivariate analysis of the effects of gender and computer vs paper/pencil modes of administration on survey results. *Dissertation Abstracts International Section A: Humanities & Social Sciences, 54*, 3247.

Moore, R. S., & Ames, G. M. (2002). Survey confidentiality vs anonymity: Young men's self-reported substance use. *Journal of Alcohol and Drug Education, 47*, 32–41.

Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology, 84*, 754–775.

Sarrazin, M. S. V., Hall, J. A., Richards, C., & Carswell, C. (2002). A Comparison of computer-based versus pencil-and-paper assessment of drug use. *Research on Social Work Practice, 12*, 669–683.

Tourangeau, R., & Smith, T. W. (1996). The impact of data collection mode, question format, and question context. *Public Opinion Quarterly, 60*, 275–304.

Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., & Sonenstein, F. L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science, 280*, 867–874.

Vereecken, C. A., & Maes, L. (2006). Comparison of a computer-administered and paper-and-pencil-administered questionnaire on health and lifestyle behaviors. *Journal of Adolescent Health, 38*, 426–432.

Walden, G. R. (2002). *Survey research methodology, 1990–1999: An annotated bibliography*. Westport, CT, US: Greenwood Press/Greenwood Publishing Group, Inc.

Wright, D. L., Aquilino, W. S., & Supple, A. J. (1998). A comparison of computer-assisted and paper-and-pencil self-administered questionnaires in a survey on smoking, alcohol, and drug use. *Public Opinion Quarterly, 62*, 331–353.