Graduate Student Journal of Psychology
2008, Vol. 10

ISSN 1088-4661

Online Surveys: Effect of Research Design Decisions on Rates of Invalid Participation and Data Credibility

Elizabeth Nosen
University of British Columbia

Sheila R. Woody
University of British Columbia

Designing online research often involves a trade-off between procedures that maximize administration efficiency and those that encourage valid and considerate participation. This article reports on three different online research design conditions (n = 413), differing in participant screening, incentive, and anonymity, that were used in a separate study on cognition in smoking cessation. High rates of invalid participation were observed in a condition in which participants participated anonymously, received a $20 incentive, and were screened for eligibility online. In addition, this condition produced significantly different data (variance, covariance, and central tendency) than the other two conditions, which involved less incentive or personal eligibility screening without participant anonymity. Removal of apparent "invalid" participants on the basis of a data screening protocol corrected some, but not all, of these differences. Results indicate that online designs offering monetary incentives should implement procedures to enhance data integrity even at the cost of increased participation barriers.

Online questionnaire-based research has become increasingly popular in recent years. In comparison to traditional paper-based questionnaires, online surveys have been hailed as easier and more efficient to administer, able to reach a broader target audience, more cost effective, and offering greater participant anonymity (Krantz & Dalal, 2000; Kraut et al., 2004; Reips, 2000, 2002b; Skitka & Sargis, 2006). Perhaps most importantly, it appears that online surveys can produce valid and trustworthy results. Several studies have found that online questionnaires show psychometric properties similar to their paper-and-pencil counterparts (Buchanan, Johnson, & Goldberg, 2005; Buchanan & Smith, 1999; Meyerson & Tryon, 2003; Pettit, 2002) and that online administration can produce results comparable to paper-based questionnaires (Denscombe, 2006; Lozar Manfreda & Vehovar, 2002; McCabe, 2004; McCabe, Couper, Cranford, & Boyd, 2006; McGraw, Tew, & Williams, 2000; Smith & Leigh, 1997).

Despite these encouraging findings, concern about the validity of web-based research remains. Due to the lack of researcher control over testing environments and increased anonymity afforded to online participants, there is a greater risk that some participants will intentionally misrepresent themselves, take part in a study more than once, or provide careless, insincere responses (Birnbaum, 2004; Johnson, 2005; Konstan, Rosser, Ross, Stanton, & Edwards, 2005; Reips, 2000, 2002a, 2002b; Skirtka & Sargis, 2006).

High rates of invalid participation may jeopardize the reliability and validity of findings from online research. In an online study of risky sexual behaviour, for example, Konstan and colleagues (2005) discovered that 11% of their sample was invalid due to study ineligibility or duplicate submission. Most importantly, they found that inclusion of these cases actually produced an erroneous rejection of their null hypothesis. Fortunately, researchers have suggested several strategies to both prevent and detect invalid participation (Johnson, 2005; Konstan et al., 2005; Reips, 2000, 2002a, 2002b).

Correspondence: Elizabeth Nosen, University of British Columbia, Department of Psychology, 2136 West Mal
Vancouver, B.C. Canada V6T 1Z4, Email: lnosen@psych.ubc.ca

Preventing Invalid Participation

Reducing motivation. Individuals may be more likely to participate repeatedly, or to respond carelessly or disingenuously, when a strong tangible incentive is offered. Konstan et al. (2005), for example, offered $20 for participation in their web-based study and had one individual provide at least 65 unique survey submissions. Malicious or mischievous responses may be less likely when individuals are personally invested in providing valid answers, such as when individuals are provided personalized feedback based on their answers to study questions (Fraley, 2004; Johnson, 2005). Not all research, however, readily lends itself to such feedback. In lieu of providing a personal growth experience, more tangible incentives may be required to ensure high survey completion rates. Two recent meta-analyses showed that offering survey participants an entry into a lottery for money or prizes significantly increased both survey response and retention rates (Göritz, 2006). That said, completion rates can be questionably low even with lottery incentives (Bosnjak & Tuten, 2003; Frick, Bächtiger, & Reips, 2001; Marcus, Bosnjak, Lindner, Pilischenko, & Schutz, 2007; O'Neil &

Penrod, 2001; O'Neil, Penrod, & Bornstein, 2003). On balance, does offering a substantial incentive promote good response and retention while providing valid data in online surveys? Further research is required to answer this question and determine the extent to which cases like those reported by Konstan et al. (2005) reflect isolated occurrences or a more ubiquitous threat to integrity of online research.

Reducing capacity. Researchers have suggested several methods for limiting capacity for invalid participation. Among these are informing participants about the detrimental effects of invalid participation, disallowing submissions with duplicate identifying information (e.g., IP addresses and cookies) and collecting personally identifying information (Reips, 2000, 2002b; Reips, Eid, & Diener, 2006). Because simply asking for personal information does not guarantee online participants will provide it, connecting this request to honorarium payment procedures may be helpful. Incentive payment methods that require only an email address (e.g., gift certificate, PayPal) allow participants to remain relatively anonymous if they wish, but, given the ease of acquiring multiple email addresses, are unlikely to reduce the risk of repeat respondents. Payment by check, on the other hand, essentially guarantees collection of valid names and mailing addresses. Personal pre-survey contact with participants, such as personalized participation invitations, can enhance web survey completion rates (Heerwegh, Vanhove, Matthijs, & Loosveldt, 2005) and may serve as an additional protection against invalid participation. Personal pre-survey screening of participants (e.g., to determine study eligibility) may discourage repeat survey submission or careless, random responding because it reminds participants of the real, non-virtual aspect of the project, provides participants an opportunity to ask questions, and increases participation burden (which may in turn increase personal investment in the project).

While these design strategies sound like common sense, many of them carry a cost, so it is important to investigate empirically whether they make a difference in rates of invalid participation and whether such procedural barriers effectively offset the risks raised by meaningful monetary incentives.

Detecting Invalid Participation

Identifying and removing invalid participants prior to data analysis is another strategy for enhancing integrity of online research. Common strategies include searching cases for matching case identifiers (e.g., email or IP addresses), examining data for response patterns indicating duplicate participation or inattentiveness, cross-validating participant eligibility criteria, and looking for suspiciously short completion times (Gosling, Vazire, Srivastava, & John, 2004; Johnson, 2005; Konstan et al., 2005; Reips, 2000). To the extent that these strategies effectively detect cases of repeat participation, deliberate misrepresentation or careless responding, their use may mitigate the risks created by high levels of invalid participation.

Little is known, however, about the effectiveness of these procedures. One of the most important issues concerns the extent to which researchers can rely on these strategies to preserve the integrity of their data in the face of high rates of invalid participation. In an ideal world, removing cases identified as invalid by these screening measures should leave data that is trustworthy and useable. In other words, these strategies should ensure that the data obtained under high-risk designs (e.g., when providing substantial monetary incentives with few procedural barriers to invalid participation) looks similar to that acquired under lower risk designs (e.g., when motivation or capacity for invalid participation is restricted). To the extent that removing cases identified as invalid by these protocols does not correct between-sample differences in data, it suggests one of two things: either these techniques are not catching all cases of invalid participation or differences in study designs are attracting fundamentally different samples. Either way, such a result would suggest that such strategies cannot be relied upon to preserve a valid sample.

Present Study

The opportunity to gain insight into several of these issues arose while collecting data for an online survey of individuals who were attempting to quit smoking (Nosen & Woody, submitted). Over the course of data collection, we encountered several problems (i.e., poor participant retention and high rates of invalid participation) that necessitated alteration of the study design. In the end, data were obtained in three phases differing in motivation and capacity for invalid participation; one at high risk (strong incentive and weak barriers to fraudulent participation) and two at lower risk (weak incentive and weak barriers; strong incentive and strong barriers). The present study examines how these design conditions differ in rates of participant recruitment, retention, and invalid participation. We also examined how the data produced by high-risk designs compared with those produced by low-risk designs on the same measures and whether removing the cases identified as invalid by available post-hoc data screening techniques effectively corrected these differences.

Because participants were not randomly assigned to conditions, this was not a true experiment. That said, it is unlikely that individuals truly interested in participating fraudulently will volunteer to take part under any condition other than the one from which they feel they can benefit. This means that investigating how these types of design decisions affect rates of invalid participation is not amenable to random assignment. The present study therefore represents a naturalistic study of the predictors and consequences of fraudulent participation in online survey research.

Method

Participants

Participants were English-speaking adults currently engaged in a serious effort to quit smoking. Participants were eligible if they reported having been a regular smoker for more than a year prior to recently beginning an attempt to quit. A total of 413 individuals completed the survey and asked us to use their data. Average reported age was 34.3 years (*SD* = 10.1). Across conditions, most participants indicated they were female (64.2%), Caucasian (83.5%), and living in Canada, the United States or the United Kingdom (96.6%). Most participants reported being employed full-time (53.0%) and having completed a post-secondary degree (47.5%). Table 1 provides details of participants' reported demographic and smoking history backgrounds broken down by design condition.

Procedure

Design. Data were obtained in three phases varying in the degree to which they provided a monetary incentive and protected against invalid participation. In the first phase of data collection (condition *A; n* = 100), participants were entered into a draw for a $200 gift certificate (i.e., low incentive). Because of the low incentive to participate, minimal barriers to prevent invalid participation were implemented; participation was impersonal (no researcher contact) and could be anonymous (only valid email required). After three months of collecting data in condition *A*, we decided to introduce an incentive to speed data collection and ensure adequate survey completion rates (see results for further details); this created a second design condition.

Participants in the second phase of data collection (condition *B; n* = 203) chose between a $20 check or a $20 online gift certificate as incentive (i.e., high incentive). In the interests of maintaining congruence with data collected in condition *A*, procedures were kept as similar as possible. As such, minimal barriers to prevent invalid participation were implemented in condition *B*—participation was again impersonal (no researcher contact) and could be anonymous (only valid email requested). After a few days of collecting data in condition *B*, suspicions of invalid participation (see results for further details) prompted us to implement greater procedural barriers; this created the third design condition.

In the final phase of data collection (condition *C; n* = 110), participants received a $20 check and were personally screened by researchers via email or telephone. Condition *C* thus involved high incentive to participate in combination with strong procedural barriers to invalid participation, as participants were not anonymous and engaged in personal contact with researchers before gaining access to the online survey. To summarize, conditions *A* and *B* used the same impersonal and anonymous procedure but differed in incentive. Conditions *B* and *C* had the same $20 incentive, but the screening procedure differed in the degree of personal communication (and anonymity). Conditions *A* and *C* were different in both incentive and screening procedures. All procedures were approved by the behavioral research ethics board at our university.

Recruitment. Participants were sequentially recruited under condition *A*, then *B*, then *C*. Recruitment was similar for all conditions; researchers placed links on smoking cessation websites and online discussion forums and posted advertisements in transit stations, universities, hospitals and health centers in the Vancouver area.

Eligibility screening. All participants were screened for study eligibility prior to participation using identical questions about age, language, and smoking history. Screening questions were asked online, immediately before completion of the survey (conditions *A* and *B*), or personally via email or phone (condition *C*). Participants who were screened online (conditions *A* and *B*) took part by clicking on a direct link to the study; qualified participants based on responses to multiple-choice questions were automatically directed to the informed consent page of the survey. Participants who were personally screened (condition *C*) were instructed to call or email the investigators for more information and to participate. Via telephone or email, research assistants screened potential participants who contacted the lab using the same questions as in the online screening procedure. Eligible condition *C* participants provided their first and last names, email address, telephone number, and mailing address (all information required to receive payment in check form) and were subsequently emailed a link to the study.

Survey design. Participants completed all questionnaires over the Internet, facilitated by SurveyMonkey.com, a web-based survey development and hosting service. Informal pilot testing was conducted with volunteers from the university and community to ensure survey functionality across browsers and operating platforms. Pilot testing also addressed issues of English language proficiency and item clarity and applicability.

Informed consent. The informed consent page described the purpose of the study, participation requirements, and confidentiality. Participants were informed that questionnaire responses would be stored in a secure, encrypted database, that all answers would remain confidential, and that data would be presented in aggregate format only. Participants were instructed that they could complete the survey as many times as they liked, but that they would receive remuneration one time only. Before completing the questionnaires, participants were required to indicate that they had one hour available to complete the survey. Participants who were screened online (conditions *A* and *B*) were asked to provide an email address for study purposes and a home mailing address as a back-up method of contacting participants for remuneration. Names were optional.

Questionnaire presentation. Each questionnaire occupied its own page, with questionnaire-specific instructions provided at the top of each page. Most questions were multiple choice, involving clicking on a box displayed below the

response choices; a few questions required a typed response. Participants continued through the survey by clicking a "next" button at the bottom of each page. Missing responses were not permitted, and respondents were not allowed to return to a previous page after leaving it. A heading at the top of each page informed participants how far they had progressed through the survey (e.g., page 6/11). The survey consisted of 17 pages, plus two additional pages for individuals who were screened for eligibility online.

Participant opt-out. Upon completion of all the questionnaires, a question asked participants if they would like their data to be used in study analyses. This question informed participants of the importance of using valid data for the research and asked participants to indicate *not* to use their data if they thought, for any reason, that their answers did not accurately reflect their true opinions (e.g., did not actually read the questions, answered randomly, filled it out pretending to be someone else). This page informed participants that they would receive remuneration for completion of the study thus far, regardless of whether they indicated that their data should be used. Following this question, a debriefing page appeared that described the study and listed some common smoking cessation resources.

Measures

Full information about the measures and the purposes of the smoking cessation cognition study is available in Nosen and Woody (submitted). Four questionnaires assessed obsessional thinking (OC cognitions): the Obsessional Beliefs Questionnaire (Obsessive Compulsive Cognitions Working Group (OCCWG), 1997, 2001, 2003), the Thought–Action Fusion Scale – Revised (Shafran, Thordarson, & Rachman, 1996), the White Bear Suppression Inventory (Wegner & Zanakos, 1994), and the Punishment scale from the Thought Control Questionnaire (Wells & Davies, 1994).

Seven measures assessed variables relevant to nicotine addiction and smoking cessation difficulty, including the Obsessive Compulsive Drinking Scale-Revised, Smoking Version (Morgan, Morgenstern, Blanchard, Labouvie, & Bux, 2004), the Smoking Self-Efficacy Questionnaire (Etter, Bergman, Humair, & Perneger, 2000), the positive scale from the Smoking Effects Questionnaire (Rohsenow et al., 2003), the Center for Epidemiological Studies Depression Scale–Short Form (Kohout, Berkman, Evans, & Cornoni-Huntley, 1993), the Fagerström Test for Nicotine Dependence (Heatherton, Kozlowski, Frecker, & Fagerström, 1991), the Appraisals of Cravings Questionnaire (ACQ; Nosen & Woody, submitted), and the Catastrophic Appraisals Index (Nosen & Woody, submitted).

Post-Hoc Validity Screening Procedures

We applied several validity-screening methods suggested by other authors (Gosling et al., 2004; Johnson, 2005; Konstan et al., 2005; Reips, 2000, 2002b). Specifically, we screened cases for repeat participation, study ineligibility, and suspicious participation behavior. We did not examine for long strings of identical responses (Johnston, 2005) because individual questionnaires were relatively short (ranging from 10 to 44 items). Nor did we examine for nearly identical survey submissions (as would occur when someone pressed the "back" button and changed only a few answers to make their submission appear new; Johnston, 2005) because survey software did not permit participants to return to a previous page.

Repeat participation. Participants were identified as repeat respondents if they provided information matching that of a previous participant for either 1) the preliminary portion of an email address, 2) a name and home address, or 3) an IP address. Duplicate cases were identified using the LAG function in SPSS.

Cross-check of eligibility criteria. Eligibility for the study was cross-checked with information participants provided in the screen with information provided later in the survey. Variables checked were participants' age, duration of regular smoking before quitting, time elapsed since quitting smoking, and number of cigarettes smoked before quitting.

Suspicious participation behavior. Two methods were used to detect individuals responding in a deceitful, careless, or inattentive fashion. First, we searched for cases that appeared to gain access to the survey by responding to the online screening questions in a trial-and-error fashion. These cases were identified as participants who completed the questionnaires within one hour of a case with the same IP address that had previously answered the screening questions in an ineligible fashion.

We also identified cases with suspiciously short survey completion times. Long completion times were not used as an exclusionary criterion because the survey program did not record completion time for individual pages. As such, long completion times are ambiguous because they could reflect something relatively innocuous, such as a person opening the first information page of the survey then deciding they would rather complete it later, or something more questionable, such as someone filling out half of the survey at one time and the other half at a later date.

To identify a cut-point for unacceptably short completion times, we examined how long trustworthy respondents took to finish the survey. During pilot testing, individuals who were known to the experimenter, highly educated, and very familiar with the questionnaires took approximately 35-40 minutes to complete the full survey while skimming questions and selecting random answers (to test the pro gramming). Participants in the conditions hypothesized to be at lower risk for invalid participation (condition*s A* and *C*) completed the survey in an average of 44.9 minutes (SD = 15.8)[1]; fewer than 5% of these participants completed the

[1] Not including participants with abnormally long completion times (i.e., over two hours, n = 24), which could represent, for example, taking a break from the survey before completing it.

Table 1
Demographic and Smoking History Variables Before and After Validity Screening

	Full Sample			After Removing "Invalid" Cases		
Variable	Condition *A* $n = 100$	Condition *B* $n = 198$	Condition *C* $n = 110$	Condition *A* $n = 89$	Condition *B* $n = 83$	Condition *C* $n = 92$
Gender (% female)	64.0% [ab]	57.6% [a]	76.4% [b]	61.8% [a]	67.5% [a]	78.3% [a]
Ethnicity (% Caucasian/European)	95.0% [a]	73.9% [b]	90.9% [a]	96.6% [a]	85.5% [b]	92.4% [ab]
Employment (% working)	87.0% [a]	86.2% [a]	70.9% [b]	86.5% [a]	79.5% [a]	69.6% [b]
Age	37.86 (9.78) [a]	28.00 (7.41) [b]	36.50 (11.30) [a]	38.60 (9.74) [a]	31.79 (8.12) [b]	39.67 (10.87) [a]
Years of Education	14.71 (2.61) [a]	14.79 (2.29) [a]	14.49 (2.43) [a]	14.67 (2.68) [a]	14.27 (2.07) [a]	14.63 (2.47) [a]
Cigarettes per day, before quit	22.88 (9.80) [a]	18.64 (9.86) [b]	20.24 (7.92) [ab]	23.35 (10.06) [a]	20.52 (7.66) [a]	20.79 (7.76) [a]
Years smoking, before quit	18.25 (10.59) [a]	9.01 (6.89) [b]	20.03 (11.91) [a]	19.44 (10.35) [a]	12.00 (7.46) [b]	21.21 (11.53) [a]
Previous quit attempts	6.34 (7.19) [a]	5.48 (7.16) [a]	6.48 (10.77) [a]	6.35 (7.56) [a]	5.88 (9.20) [a]	6.92 (11.66) [a]
Longest previous attempt (months)	8.55 (12.85) [a]	6.61 (12.06) [a]	7.99 (14.98) [a]	8.84 (13.88) [a]	5.40 (7.94) [a]	10.03 (26.68) [a]
Months into current quit attempt	1.74 (6.49) [a]	5.75 (10.74) [b]	2.26 (5.84) [c]	0.82 (1.02) [a]	2.03 (1.74) [b]	1.43 (1.45) [b]
Reduction in cigarettes per day	21.78 (10.49) [a]	16.73 (10.11) [b]	18.54 (8.39) [ab]	22.52 (10.65) [a]	18.10 (8.45) [b]	19.77 (7.80) [ab]
Cigarettes per day, at time of assessment	1.10 (3.46) [a]	1.69 (5.18) [a]	1.90 (3.82) [a]	0.81 (2.36) [a]	2.42 (4.02) [a]	1.01 (2.91) [a]

Note: Disimilar superscripts ([a, b, c]) indicate that the percentages or means (shown with standard deviations) associated with the design condition are significantly different, based on either Chi-square or ANOVAs with post-hoc Fisher's LSD tests, $p > .05$.

survey in less than 25 minutes. Thus, completion times of less than 25 minutes were taken as an indicator of participants who were unlikely to have fully read or adequately considered all of the questions.

Results

Participant Recruitment and Retention

In condition *A*, 156 participants completed the first page of the survey over a 10-week period, for an average of one interested and willing participant about every 11 hours. Of these, 102 completed the survey, and 100 asked that we use their data (a 64.1% retention rate). Under condition *B*, 238 people completed the first page of the survey over a 54-hour period (one willing participant about every 14 minutes). Of these, 211 participants finished the survey and 203 asked that we use their data (85.3% retention rate). Finally, 115 participants in condition *C* completed the first page of the survey over a 10-week period (one willing participant about every 14 hours), of which 111 completed the survey, and 110 asked us to use their data (a 95.6% retention rate). Table 1 provides details on how the three samples differed in terms of reported demographic and smoking variables of interest to the original study.

Rates of Invalid Participation Across Groups

Table 2 presents the number of participants flagged by validity screening protocols in each condition. After using all screening protocols to flag cases with indicators of invalid participation, chi-square analyses revealed that significantly more participants in condition *B* were flagged as invalid (58.6% of survey completers) than in either condition *A* (11.0% of survey completers), χ^2 ($n = 303$) = 62.02, $p < .01$, $\phi = -.45$, or condition *C* (16.4% of survey completers), χ^2 ($n = 313$) = 51.76, $p < .01$, $\phi = -.41$. Conditions *A* and *C* did not differ in the proportion of cases identified as invalid, χ^2 ($n = 210$) = 1.26, $p > .05$, $\phi = -.08$.

Effects of Design Condition on Data

To better understand the effects of research design decisions (and essentially, invalid participation) on results obtained, we examined the questionnaire data for group differences in variance, covariance, and central tendency. We anticipated that the group participating under strong participation incentive with few procedural barriers to invalid participation (condition *B*) would differ from the other two conditions, despite the fact that our advertisement was the same and the study questionnaires were conceptually unrelated to the recruitment conditions. We did not expect the groups at lower risk for invalid participation (conditions *A* and *C*) to differ, despite being maximally different in incentive and barriers to invalid participation.
Variables were first examined for missing values, outliers, and fit with multivariate assumptions. Next, two om-nibus multivariate analyses of variance (MANOVAs) were conducted on conceptually related questionnaires; one analyzed the four-smoking-related variables, one analyzed the seven cognition variables. Between-group differences in covariance, variance, and central tendency were tested using Box's test of equality of covariance matrices, Levene's test for equality of error variances, and MANOVAs,

Table 2
Frequency of Indicators of Invalid Participation Across Conditions

Type of Invalid Participation	Design Condition		
	A (*n* = 100)	*B* (*n* = 203)	*C* (*n* = 110)
Repeat participation			
Duplicate IP address	3 (3.0%)	26 (12.8%)	2 (1.8%)*
Duplicate identity (name and address)	0	0	0
Duplicate email address†	0	1 (0.5%)	0
Ineligible participants			
Accessed survey through trial and error	1 (1.0%)	34 (16.7%)	n/a
Failed eligibility cross-check	6 (6.0%)	56 (27.6%)	11 (10.0%)
Short completion time (< 25 minuntes)	4 (4.0%)	82 (40.4%)	5 (4.5%)
Total invalid cases among survey completers	11 (11.0%)	119 (58.6%)	18 (16.4%)

Note: Percentages reflect proportion of invalid cases among individuals who completed the survey and asked that we use their data.
† Matching portion of email address before the "@" symbol
* Because participants in condition *C* had provided reliable names and mailing addresses, further examination of the database permitted confidence that these two duplicate IP addresses represented unique individuals sharing a household with another participant.

respectively. Omnibus gateway analyses were used where available and alpha was set at .05 for all tests.[2]

Through a technical error in the computer administration of the survey, three cases from condition *B* were missing all responses on one questionnaire each (affecting the WBSI, OCDS, and SSEQ). These cases were removed from analyses. No other cases had missing data. Univariate outliers were replaced with scores adjacent to the next highest or lowest. This procedure affected five scores on the OBQ (three from condition *B*, two from condition *C*), one score on the TCQ punishment subscale (from condition *C*), and two scores on the WBSI (from condition *B*). Scores on the Catastrophic Appraisals Index (CAI) were moderately positively skewed for conditions *A* and C, but were severely negatively skewed for condition *B*. As such, the CAI was excluded from analyses of variance and covariance but was included in analyses of central tendency. After removing two multivariate outliers from condition *B*, 408 cases were available for analyses (100 in condition *A*, 198 in condition *B*, and 110 in condition *C*).

Covariance. Table 3 provides details of Box's tests of equality of covariance matrices. When using the four cognition questionnaires as dependent variables, Box's test was significant for two of the three between-group comparative analyses, indicating that the covariance matrix of condition *B* was significantly different from the covariance matrices produced by both condition *A* and condition *C*. Box's test was not significant when conditions *A* and *C* were compared, indicating similar patterns of variable interrelationships in these groups.

For the smoking-relevant questionnaires, Box's test was significant for one comparison, indicating that the covariance matrix of condition *A* was significantly different from the covariance matrix of condition *B*. Box's test suggested that the pattern of variable inter-relationships in condition *C* was not significantly different from that of condition *A* or *B*.

Variance. Standard deviations can be found in Table 4. Levene's test for equality of error variances was conducted to investigate between-group differences in data variability. As an omnibus test, Levene's test was significant for five of the 10 scales, indicating that the variability differed in some way between the three groups for the TCQ-punishment, WBSI, TAFS, SSEQ, and ACQ, *F's* (2, 416) ≥ 3.79, .06 ≤ |*r's*| ≤ .07, $p < .05$. Levene's test showed homogeneity of variance across conditions for the remaining measures, *F*'s (2, 416) ≤ 2.89, |*r*'s| ≤ .06, $p > .05$.

Subsequent Levene's tests comparing each of three design groups to each other indicated that, in all cases, the heterogeneity was due to anomalies in condition *B* data. Comparisons of conditions *A* and *C* indicated homogeneity of variance for all of the questionnaires, *F*'s (1, 211) ≤ 3.42, |*r*'s| ≤ .06, $p \geq .07$. Compared to condition *A*, however, condition *B* produced significantly different variability in scores on the five measures mentioned above, *F's* (1, 306) ≥ 8.69, *r*'s = .06, $p < .01$. Compared to condition *C*, condition *B* again produced significantly different variance for four of the five measures, *F's* (1, 315) ≥ 8.61, .05 ≤ |*r's*| ≤ .06, $p < .01$. Variability did not differ on WBSI across conditions *B* and *C*, $F(1, 315) = 1.57$, $r = .05$, $p > .05$.

Central Tendency. Questionnaire means are presented in Table 4. An omnibus MANOVA with the four cognition variables was statistically significant, Pillai's Trace = 0.22, $F(8, 806) = 12.22$, $\eta_p^2 = .11$, $p < .001$, indicating the presence of significant group differences in questionnaire central tendency. The omnibus MANOVA using the smoking relevant variables was also significant, Pillai's Trace = 0.13, $F(12, 802) = 4.72$, $\eta_p^2 = .07$, $p < .001$.

Follow-up analyses of variance (or Welch tests, as appropriate) revealed significant between-group differences on three of the four cognition variables and five of the six smoking-relevant measures. Significant differences were observed on all measures (*F*s > 5.39, p < .01), except the WBSI, Welch's $F(2, 204.11) = 2.74$, $r = .09$, $p > .05$, and FTND, $F(2, 405) = 0.25$, $p = .78$, $\eta_p^2 = .001$. Post-hoc Fisher's LSD and Games-Howell tests (as appropriate) suggested that the condition *B* data again drove these differences. Indeed, means in conditions *A* and *C* were not significantly different for any questionnaire, *d's* ≤ 0.21, $p > .05$. Participants in condition *B,* however*,* displayed means

[2] We opted for less stringent control of Type 1 error rates because we were more interested in examining the broad pattern of results than the significance of any one particular test.

Table 3
Box's Tests of Equality of Covariance Before and After Validity Screening

	Condition Comparisons							
	Omnibus *A* vs. *B* vs. *C*		*A* vs. *B*		*B* vs. *C*		*A* vs. *C*	
Full Sample	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Four cognition variables	5.49	<.001	7.00	<.001	5.91	<.001	0.88	>.05
Six smoking variables	1.43	<.05	2.15	<.01	1.39	>.05	0.65	>.05
"Invalid" Participants Deleted								
Four cognition variables	3.50	<.01	3.50	<.01	3.50	<.01	0.48	>.05
Six smoking variables	0.74	>.05	n/a	n/a	n/a	n/a	n/a	n/a

Notes: Full sample *N*'s for conditions *A*, *B* and *C* were 100, 198 and 110, respectively. After removing invalid participants, sample sizes were reduced to 89, 83 and 92, respectively.

that were significantly different from those in both conditions *A* and *C* on all three cognition measures that showed differences in the omnibus test and on four of the six smoking-related questionnaires showing omnibus test differences (the ACQ, SEQ-positive, CES-D, and SSEQ). On all of these measures, condition *B* means were in the direction of more pathological than those in the other conditions; $0.33 \leq d \leq 0.87$. Similarly, median tests revealed that CAI central tendency was equivalent for conditions *A* and *C* (medians = 3.00, χ^2 (N = 210) = 0.15, $p > .05$), but that the condition *B* group (median = 9.00) scored significantly higher than both of the other conditions, χ^2's (N = 298) ≥ 47.92, $p < .001$, $\phi \geq .16$).

Does Removing Invalid Participants Correct the Problem?

Because condition *B* contained a significantly greater proportion of participants identified as invalid than did the other two groups, we were interested in examining the extent to which removal of these cases would remedy the group *B* anomalies in data variance, covariance and central tendency. As such, the above analyses were repeated without participants identified as "invalid" by the previously described data screening protocols. For these analyses, 264 cases were available (89 in condition *A*, 83 in condition *B*, and 92 in condition *C)*.

Covariance. Table 3 shows details of Box's test of equality of covariance matrices after removing "invalid" participants. As with the full samples, the covariance matrix of cognition questionnaires in condition *B* differed significantly from the covariance matrices of both conditions *A* and *C*, which did not differ from each other. Thus, removing participants identified as invalid by our screening procedures did not eliminate group differences in covariance among the cognition variables.

The smoking-relevant questionnaires showed a different result. Removing participants who were identified as invalid using the data screening procedures resulted in equivalent covariance matrices across groups.

Variance. Standard deviations for questionnaires after removing "invalid" participants can be found in Table 4. Removing participants identified as invalid reduced the number of variables for which there was cross-condition heterogeneity in variance from five to three questionnaires: TCQ-punishment, TAFS, and CES-D, *F'*s (2, 416) ≥ 3.29, *r*'s = .08 - .09, $p < .05$. Follow-up tests showed that condition *A* was discrepant from the other two conditions for CES-D variance, *F'*s (1, 416) ≥ 4.49, *r*'s = .07, *p*'s < .05. As with the full sample, Condition *B* was discrepant from the other two conditions on the TCQ-punishment and TAFS, *F'*s (2, 416) ≥ 5.06, *r*'s = .07, $p < .05$. Thus, removing "invalid" participants corrected some, but not all, of the between-group differences in questionnaire variability.

Central Tendency. Repeating the MANOVAs showed that group differences in central tendency still remained in both the cognition variables and the smoking variables after removing invalid participants, Pillai's Trace's ≥ 0.11, *F'*s ≥ 2.58, $p < .01$, $\eta_p^2 \geq 0.06$. Follow-up tests showed that all differences on cognition measures that had been apparent in the full sample were still present following removal of invalid participants. In all cases (with the OBQ, TCQ-punishment, and TAFS) condition *B* means were significantly different, $d \geq 0.36$, $p < .05$, from those of conditions *A* and *C,* which did not differ, $d < 0.18$, $p > .05$.

Some of the differences observed on smoking relevant variables in the full sample were ameliorated by removing invalid participants, but three of the six variables still showed differences; the SSEQ and ACQ, *F*'s (2, 261) ≥ 7.50, $p < .01$, as well as the CES-D, Welch's *F* 172.35) = 5.17, $p < .01$. As before, condition *B* was discrepant from the other two conditions, $d \geq 0.45$, $p < .05$, which did not differ, $d < 0.18$, $p > .05$. Median tests on the CAI were also unchanged from the full sample results, with participants in condition B (median = 7.00) scoring significantly higher than participants in conditions A and C (medians = 3.00), χ^2 's (n's ≥ 172) ≥ 6.08, $p < .05$, _ ≥ .03. In sum, many between-group differences in central tendency remained after removing "invalid" participants.

Table 4
Means and Standard Deviations of Study Questionnaires Before and After Validity Screening

	Condition *A*		Condition *B*		Condition *C*	
	M	*SD*	*M*	*SD*	*M*	*SD*
Full Sample	*n* = 100		*n* = 198		*n* = 110	
Craving Severity (OCDS)	25.40	6.30	26.89	6.22	24.52	6.52
Cessation Self-Efficacy (SSEQ)	38.40	12.81	**34.82**	**9.71**	40.16	12.12
Positive Smoking Expectations	1.28	0.56	**1.44**	0.58	1.25	.52
Depression (CES-D)	8.33	4.63	**9.83**	4.19	7.63	4.24
Nicotine Dependence (FTND)	5.46	2.15	5.30	2.16	5.45	2.50
Appraisals of Cravings (ACQ)	72.60	37.99	**91.34**	**30.53**	68.01	36.53
Thought Suppression (WBSI)	50.30	11.89	53.42	9.43	52.02	11.44
Obsessional Beliefs (OBQ)	146.00	37.18	**173.61**	41.74	138.40	38.45
TCQ-punishment	9.26	2.20	**12.11**	**4.11**	9.01	2.40
Thought Action Fusion (TAFS)	18.84	13.03	**29.78**	**18.22**	17.74	12.77
"Invalid" Participants Deleted	*n* = 89		*n* = 83		*n* = 92	
Craving Severity (OCDS)	25.49	6.52	27.27	7.08	25.08	6.68
Cessation Self-Efficacy (SSEQ)	39.26	12.76	**33.82**	11.22	40.58	12.24
Positive Smoking Expectations	1.29	0.58	1.35	0.60	1.30	0.52
Depression (CES-D)	8.19	**4.77**	9.67	4.02	7.79	3.95
Nicotine Dependence (FTND)	5.51	2.13	5.49	2.29	5.85	2.35
Appraisals of Cravings (ACQ)	74.48	37.54	**90.63**	33.09	67.84	36.39
Thought Suppression (WBSI)	50.55	12.00	54.35	10.30	52.75	10.97
Obsessional Beliefs (OBQ)	145.48	37.12	**165.07**	40.42	138.47	39.35
TCQ-punishment	9.29	2.11	**10.90**	**3.56**	9.04	2.30
Thought Action Fusion (TAFS)	18.51	12.72	23.63	**15.69**	17.41	12.58

Note: Bold figures (means or standard deviations) are significantly different from comparable figures in the other two conditions. Underlined figures indicate significant difference from one other condition. TCQ-punishment refers to the punishment subscale from the Thought Control Questionnaire.

Discussion

Results of the current study suggest that online survey design decisions have important implications for the reliability and validity of research results. When few procedural barriers to invalid participation were present, provision of a strong monetary incentive (condition *B*) was associated with extremely high rates of questionable survey submissions, including cases of repeat participation, study ineligibility, and improbably fast completion times. Rates of invalid participation were significantly lower when either the incentive was reduced (while limited procedural barriers were maintained, as in condition *A*) or when procedural barriers in the form of personal participant contact and limited anonymity were implemented (while the incentive was maintained, as in condition *C*). Unsurprisingly, condition *B* produced data that were significantly different, in terms of questionnaire covariance, variance, and central tendencies, from the data obtained in the other two conditions. Deletion of participants identified through data screening procedures as "invalid" corrected some, but not all, of these differences in the data.

How Design Affects Ease of Data Collection

The design features of condition *B* (high monetary incentive and anonymous, automated participation) had one big advantage, which was ease of participant recruitment. With a seemingly eligible and willing participant completing the study every 14 minutes, we reached our targeted sample size in only three days of data collection. Recruitment was much slower in the conditions lacking the high monetary incentive (condition *A*) or implementing greater participation barriers (condition *C*); both of these conditions averaged about two new willing and eligible participants per day.

Interestingly, recruitment rates were similar for conditions *A* and *C*, despite the fact that condition *C* involved a guaranteed financial payoff, whereas the incentive in condition *A* was an entry in a drawing. The personal screening and decreased anonymity in condition *C* likely contributed to this effect. Contacting the researcher, waiting for a response, providing answers to the screening questions, and waiting for a decision about eligibility are much more burdensome for the participant than simply clicking on a link and beginning to complete the survey. Also, in the era of identity theft and online fraud, providing one's *real* name, date of birth, email address, home address, and phone num-

ber to someone who claims to be a researcher in Canada may be a risk many people would rather not take. Despite the increased burden, however, offering a monetary incentive under more stringent participant screening conditions did seem to bolster participant retention, which was almost 96% in condition *C*, markedly higher than in condition *A* (65%).

How Design Affects Risk of Invalid Participation

Invalid participation, which we judged on the basis of data screening procedures indicating repeat participation, accessing the survey through trial and error, failure to meet study eligibility criteria, and improbably fast survey completion times, was detected in condition *B* at rates nearly five times those seen in the other two conditions.

Variations in condition incentive, anonymity, and personal contact likely contributed to this difference by influencing both motivation and capacity for fraudulent participation. In particular, participants in conditions *B* and *C* (but not in *A*) had financial motivation for fraudulent participation. At the same time, participants in conditions *B* and *A* (but not in *C*) had few procedural barriers to invalid participation. In other words, participants in condition *B* had both motivation and capacity for fraudulent participation, while participants in the other two conditions were missing one of these critical elements. The fact that no participant in condition *B* provided an email address matching that of another participant, despite many obvious instances of repeat survey submission, seems to support the idea that many of these participants deliberately took part under false pretenses.

Research suggests that offering tangible rewards for behavior may undermine intrinsic motivations (Deci, Koestner, & Ryan, 1999). Thus, participants in conditions *B* and *C* may have been particularly likely to have been motivated by the monetary reward rather than, for example, the pleasure of helping science or learning more about themselves or the research. In and of itself, this may not be disastrous. The problem, rather, may lie in the fact that participation in condition *B* was anonymous and automated, such that it required little effort or personal investment from participants. Termed the "overjustification effect" (Lepper, Greene, & Nisbett, 1973), individuals offered a larger payment than necessary may perceive the task more negatively (e.g., less enjoyable) than tasks associated with lower incentives (Freedman, Cunningham, & Krismer, 1992). Anonymity and lack of personal contact with the researchers may have compounded the effects of poor motivation and negative expectations by removing social pressures for conscientious behaviour.

How Design Affects Integrity of Collected Data

Practically speaking, the more important questions here concern data integrity in the context of high rates of invalid participation. With respect to similarity of the data obtained under different incentive and anonymity conditions, the data produced by condition *B* was anomalous from the other two conditions in questionnaire variance, covariance, and central tendency. Highlighting the irregular data obtained in condition *B*, data did not differ in the conditions at low risk for invalid participation (conditions *A* and *C*). The types of data abnormalities seen in condition *B* would critically threaten the validity of any findings based on the sample. Not only would they complicate analyses by violating statistical assumptions and distorting Type I error rates, but also they would raise serious questions about the representativeness of the sample.

The most parsimonious explanation for the anomalous data in condition *B* is that invalid participation wreaked havoc on the data. Indeed, it is easy to imagine how high rates of repeat participation, misrepresenting study eligibility, or failing to devote adequate care and attention to the task could affect results. The possibility exists, then, that simply removing the questionable individuals from the sample could salvage the data. If these procedures can effectively remedy the data abnormalities produced by using a high-risk design, then reducing the incentive or implementing burdensome barriers to participation may not be necessary to ensure trustworthy data.

In the present study, removing participants who were flagged by validity screening protocols corrected many, but not all, between-group differences in questionnaire variability, central tendency, and covariance. This suggests that either the screening techniques used were not 100% effective at identifying invalid participants, or that there were fundamental differences in the samples attracted by the various design conditions.

In support of the former explanation, most post-hoc validity screening strategies are subject to particular weaknesses. Participants could avoid submitting surveys with duplicate IP addresses, for example, by taking part on different computers at a library or another public Internet access site. Detecting individuals who do not genuinely meet study eligibility criteria is also fallible. In particular, many individuals could have guessed the "correct" responses to screening questions based on recruitment advertisements or information passed on from previous participants. Asking about eligibility criteria later in the survey, using different wording and response formats (e.g., open-ended vs. multiple-choice), may identify some (but probably not all) of these participants. Finally, screening for implausibly fast completion times based on total (not page-by-page) survey completion times would miss invalid participants who took their time, who actually read the questions, or who took a break mid-survey.

Data differences that remained after removing invalid participants could also be due to fundamental, legitimate differences in the samples attracted to participate under the various design conditions. Indeed, individuals displaying more psychological difficulties (e.g., more pathological obsessive compulsive or smoking-related thoughts and behaviors) may be less likely to take part in online studies

that are not both highly rewarding and relatively painless. This possibility is consistent with evidence indicating that survey design features like method of informed consent and amount of personal information requested can influence sample characteristics (O'Neil et al., 2003). Design choices related to incentive and barriers to invalid participation may similarly affect sample representativeness.

Limitations

Several factors are important to bear in mind when considering the results of the current study. First, participants were recruited sequentially (into condition A, then *B*, then *C*), without random assignment. Thus, we cannot be certain that the anomalies seen in condition *B* are due solely to differences in design. Nevertheless, this remains the most parsimonious explanation. Additionally, the nature of this topic does not readily lend itself to more conclusive experimental designs. In particular, individuals intending to profit by fraudulent participation are unlikely to take part if randomly assigned to incentive. Another caveat to note concerns the post-hoc validity screening strategies. Those employed were tailored to the current study, as they would be in other forms of online research. However, this means that some potentially useful strategies were not applied. For example, we did not exclude on the basis of long strings of identical responses because individual questionnaires were relatively short. Thus, use of different strategies could potentially yield different rates of invalid case detection.

Recommendations

Results suggest that providing a sizeable monetary incentive to online participants is an acceptable strategy to boost recruitment and retention, as long as strong procedural barriers to invalid participation are implemented. Apart from an improved retention rate, this design did not convey apparent advantages over offering a minimal incentive with fewer procedural barriers. If retention rate is not a concern, then a design offering minimal incentive and anonymous, automated participation may be the cheapest and easiest solution. Either way, researchers should take measures to ensure that participants do not have both motivation and capacity to profit from participation. Use of data validity screening techniques is also strongly recommended, particularly since we found rates of invalid participation as high as 16.4% among designs at *low* risk for fraudulent participation. That being said, current results suggest that use of these validity screening strategies does not fully ameliorate the problem of invalid participation.

References

Birnbaum, M. H. (2004). Human research and data collection via the Internet. *Annual Review of Psychology, 55*, 803-832.

Bosnjak, M., & Tuten, T. L. (2003). Prepaid and promised incentives in web surveys: An experiment. *Social Science Computer Review, 21*(2), 208-217.

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*(2), 115-127.

Buchanan, T., & Smith, J. L. (1999). Research on the Internet: Validation of a World-Wide Web mediated personality scale. *Behavior Research Methods, Instruments & Computers, 31*(4), 565-571.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627-668.

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*(2), 246-254.

Etter, J.-F., Bergman, M. M., Humair, J. P., & Perneger, T. V. (2000). Development and validation of a scale measuring self-efficacy of current and former smokers. *Addiction, 95*(6), 901-913.

Fraley, R. C. (2004). *How to conduct behavioral research over the Internet*. New York: Guilford Press.

Freedman, J. L., Cunningham, J. A., & Krismer, K. (1992). Inferred values and the reverse-incentive effect in induced compliance. *Journal of Personality and Social Psychology, 62*(3), 357-368.

Frick, A., Bächtiger, M.-T., & Reips, U.-D. (2001). Financial incentives, personal information and drop-out in online studies. In U.-D. Reips & M. Bos_njak (Eds.), *Dimensions of Internet science* (pp. 209-219). Lengerich, Germany: Pabst.

Göritz, A. S. (2006). Incentives in Web-based studies: Methodological issues and a review. *International Journal of Internet Science 1*(1), 58-70.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist, 59*(2), 93-104.

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerström, K.-O. (1991). The Fagerström Test for Nicotine Dependence: A revision of the Fagerström Tolerance Questionnaire. *British Journal of Addiction, 86*(9), 1119-1127.

Heerwegh, D., Vanhove, T., Matthijs, K., & Loosveldt, G. (2005). The effect of personalization on response rates and data quality in Web surveys. *International Journal of Social Research Methodology: Theory & Practice, 8*(2), 85-99.

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality, 39*(1), 103-129.

Kohout, F. J., Berkman, L. F., Evans, D. A., & Cornoni-Huntley, J. (1993). Two shorter forms of the CES-D Depression Symptoms Index. *Journal of Aging and Health, 5*(2), 179-193.

Konstan, J. A., Rosser, B. R. S., Ross, M. W., Stanton, J., & Edwards, W. M. (2005). The story of subject naught: A cautionary but optimistic tale of Internet survey research. *Journal of Computer-Mediated Communication, 10*(2), No Pagination Specified.

Krantz, J. H., & Dalal, R. (2000). Validity of Web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 35-60). San Diego, CA: Academic Press.

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online: Report of Board of Scientific Affairs' Advisory Group on the conduct of research on the Internet. *American Psychologist, 59*(2), 105-117.

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the 'overjustification' hypothesis. *Journal of Personality and Social Psychology, 28*(1), 129-137.

Lozar Manfreda, K., & Vehovar, V. (2002). Do mail and web surveys provide the same results? In A. Ferligov & A. Mrva (Eds.), *Development in social science methodology* (pp. 149-169). Ljubljana, Slovenia: Metodoloski zvezki.

Marcus, B., Bosnjak, M., Lindner, S., Pilischenko, S., & Schutz, A. (2007). Compensating for low topic interest and long surveys: A field experiment on nonresponse in Web surveys. *Social Science Computer Review, 25*(3), 372-383.

McCabe, S. E. (2004). Comparison of Web and mail surveys in collecting illicit drug use data: A randomized experiment. *Journal of Drug Education, 34*(1), 61-72.

McCabe, S. E., Couper, M. P., Cranford, J. A., & Boyd, C. J. (2006). Comparison of Web and mail surveys for studying secondary consequences associated with substance use: Evidence for minimal mode effects. *Addictive Behaviors, 31*(1), 162-168.

McGraw, K. O., Tew, M. D., & Williams, J. E. (2000). The integrity of Web-delivered experiments: Can you trust the data? *Psychological Science, 11*(6), 502-506.

Meyerson, P., & Tryon, W. W. (2003). Validating Internet research: A test of the psychometric equivalence of Internet and in-person samples. *Behavior Research Methods, Instruments & Computers, 35*(4), 614-620.

Morgan, T. J., Morgenstern, J., Blanchard, K. A., Labouvie, E., & Bux, D. A. (2004). Development of the OCDS-Revised: A measure of alcohol and drug urges with outpatient substance abuse clients. *Psychology of Addictive Behaviors, 18*(4), 316-321.

Nosen, E., & Woody, S. R. (submitted). Generalizability of cognitive theories of obsessions: Metacognitive processes in smoking cessation.

O'Neil, K. M., & Penrod, S. D. (2001). Methodological variables in Web-based research that may affect results: Sample type, monetary incentives, and personal information. *Behavior Research Methods, Instruments & Computers, 33*(2), 226-233.

O'Neil, K. M., Penrod, S. D., & Bornstein, B. H. (2003). Web-based research: Methodological variables' effects on dropout and sample characteristics. *Behavior Research Methods, Instruments & Computers, 35*(2), 217-226.

Obsessive Compulsive Cognitions Working Group (OCCWG). (1997). Cognitive assessment of obsessive-compulsive disorder. *Behaviour Research and Therapy, 35*(7), 667-681.

Obsessive Compulsive Cognitions Working Group (OCCWG). (2001). Development and initial validation of the Obsessive Beliefs Questionnaire and the Interpretation of Intrusions Inventory. *Behaviour Research and Therapy, 39*, 987-1006.

Obsessive Compulsive Cognitions Working Group (OCCWG). (2003). Psychometric validation of the Obsessive Beliefs Questionnaire and the Interpretation of Intrusions Inventory: Part I. *Behaviour Research and Therapy, 41*(8), 863-878.

Pettit, F. A. (2002). A comparison of world-wide web and paper-and-pencil personality questionnaires. *Behavior Research Methods, Instruments & Computers, 34*(1), 50-54.

Reips, U.-D. (2000). The Web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 89-117). Burlington, MA: Academic Press.

Reips, U.-D. (2002a). Internet-based psychological experimenting: Five dos and five don'ts. *Social Science Computer Review, 20*(3), 241-249.

Reips, U.-D. (2002b). Standards for Internet-based experimenting. *Experimental Psychology, 49*(4), 243-256.

Reips, U.-D., Eid, M., & Diener, E. (2006). Web-based methods. In *Handbook of multimethod measurement in psychology*. (pp. 73-85). Washington, DC: American Psychological Association.

Rohsenow, D. J., Abrams, D. B., Monti, P. M., Colby, S. M., Martin, R., & Niaura, R. S. (2003). The Smoking Effects Questionnaire for adult populations: Development and psychometric properties. *Addictive Behaviors, 28*(7), 1257-1270.

Shafran, R., Thordarson, D. S., & Rachman, S. (1996). Thought-action fusion in obsessive compulsive disorder. *Journal of Anxiety Disorders, 10*(5), 379-391.

Skitka, L. J., & Sargis, E. G. (2006). The Internet as psychological laboratory. *Annual Review of Psychology, 57*, 529-555.

Smith, M. A., & Leigh, B. (1997). Virtual subjects: Using the Internet as an alternative source of subjects and research environment. *Behavior Research Methods, Instruments & Computers, 29*(4), 496-505.

Wegner, D. M., & Zanakos, S. (1994). Chronic thought suppression. *Journal of Personality, 62*(4), 615-640.

Wells, A., & Davies, M. I. (1994). The Thought Control Questionnaire: A measure of individual differences in the control of unwanted thoughts. *Behaviour Research and Therapy, 32*(8), 871-878.