The high-hurdle technique put to the test: Failure to find evidence that increasing loading times enhances data quality in Web-based studies

ANJA S. GÖRITZ
University of Erlangen-Nuremberg, Nuremberg, Germany

AND

STEFAN STIEGER
Medical University of Vienna, Vienna, Austria

Two Web-based experiments examined the usefulness of artificially delaying the loading of the first page of the study. The idea pursued in this technique is to filter out less-motivated respondents through a higher respondent burden in the form of waiting time. Participants who remain in the study despite having had to wait for the first page of the study to appear on the screen are expected to be more highly motivated, and thus to produce data of higher quality. In both experiments, as expected, the longer the loading time, the lower the likelihood of people responding to the study. However, contrary to expectation, the dropout rate and quality of data were independent of the loading time. Therefore, artificially delaying the loading of the first page of the study is counterproductive.

Web-based data collection bears many advantages, which have rendered it popular—for instance, through potential access to respondents from all over the world and lower costs (Birnbaum, 2004; Göritz & Schumacher, 2000; Skitka & Sargis, 2006). However, there are also disadvantages to this method in comparison with more conventional methods of data collection. There have been attempts to develop techniques and standards to prevent or at least to alleviate these disadvantages (Reips, 2002b).

One of the challenges of Web-based relative to more traditional methods of data collection is increased dropout (O'Neil, Penrod, & Bornstein, 2003). The term *dropout* denotes participants' premature abandonment of a study before the end of the questionnaire or interview is reached. Reips (2002b) has suggested dealing with dropout in two ways: (1) reduce dropout itself or (2) reduce the undesired consequences of dropout. With regard to reducing the dropout rate, material incentives have been shown to be effective (Göritz, 2004, 2005, 2006), as well as a study's layout and design (Reips, 2000, 2002a; Wenzel, 2001) and prudent usage of technologies such as JavaScript, Flash, or Java (Schmidt, 2000; Schwarz & Reips, 2001). To alleviate the undesired consequences of dropout on data quality, Reips (2002b) suggested methods such as the warm-up technique and the high-hurdle technique.

Participants are more likely to drop out at the beginning than near the end of a questionnaire (Reips, 2002c). The reason for participants dropping out at the beginning is thought to be their decision that they do not want to take part after

having satisfied their curiosity about what the study is about and how it looks and feels. By contrast, close to the end of a questionnaire—after having invested considerable time—participants are more likely to stay committed until the very end. To make the best of this pattern, Reips has suggested placing important questions toward the end of the study—a method he has termed the *warm-up technique*.

As a second method, Reips (2002b) describes the *high-hurdle technique*. This method tries to provoke participants with a low level of motivation to drop out early in the study, with the result that only highly motivated respondents remain in the sample. These remaining participants are hoped to produce data of high quality. There are different kinds of hurdles that might be put at the beginning of a study: asking personal and sensitive information, admonitions that participants can be tracked on the basis of their IP address, artificially increasing the loading times of the first page(s) of the study (for a checklist, see Reips, 2002b, p. 249).

Mimicking the high-hurdle technique, other researchers have increased respondent burden by requiring participants to type in a code to access the study rather than providing an automatic login procedure. In Crawford, Couper, and Lamias (2001), the response rate was lower and viewing time of the questionnaire was higher with the manual login than with the automatic login, whereas the numbers of open-ended questions answered were not different. In Heerwegh and Loosveldt (2002), the response rate and viewing time did not differ between the two conditions. However, the dropout rate was lower and answers to two sensitive questions more sub-

A. S. Göritz, anja.goeritz@wiso.uni-erlangen.de

322

stantial with the manual than with the automatic login. Finally, in Heerwegh and Loosveldt (2003), the response rates did not differ among automatic, semiautomatic, and manual login procedures. However, the dropout rate was lowest in the manual, second lowest in the semiautomatic, and highest in the automatic condition. There were no significant differences in the time of viewing the study, likelihood of answering an open-ended question, number of characters in the answer to the open-ended question, and likelihood of providing substantial answers to two sensitive questions.

Thus, there is some evidence that artificially increasing respondent burden by having participants key in an access code increases data quality. However, this particular form of the high-hurdle technique might also engender a sense that data are more confidential and that each individual participant was sought after for their unique personal opinion rather than the opinion of just anybody who happens to come across the study. Thus, some of the effects found in these studies might be attributable to effects that go beyond mere differences in respondent burden. Moreover, Stieger, Reips, and Voracek (2007) have increased respondent burden by using a forced response design; that is, participants could proceed in the questionnaire only if they answered every question on a page. This hurdle increased the dropout rate and caused reactance with participants' answers to some questions.

The study at hand evaluates one variant of the high-hurdle technique that has hitherto not been empirically examined—deliberately increasing the loading time of the first page of the study (see Reips, 2002b). The idea behind this technique is that this hurdle will help by sifting respondents with a high level of motivation from those with a low level of motivation. Highly motivated participants are expected to be more willing to wait until the first page of the study appears on the screen than are people with a low level of motivation. Consequently, this artificially increased respondent burden will provoke people with low motivation to drop out at the beginning of a study. Participants who remain committed despite this hurdle are expected to produce data of higher quality than would an unsifted sample, where no such hurdle was present.

Two studies were conducted to examine the usefulness of artificially delaying the loading of the first questionnaire page. We take into account four aspects: (1) Does the application of this variant of the high-hurdle technique fulfill its stated aim of sifting out people? That is, does the likelihood of responding truly decrease as loading time increases? (2) Does the application of this technique fulfill its stated aim of reducing dropout after the hurdle? (3) Because a respondent's staying in the study despite the hurdle does not necessarily need to be caused by his or her high level of motivation, we tested for several indicators of motivation—namely, whether participants who jump the hurdle skip fewer items, are more likely to answer open-ended questions, and type in longer answers to open-ended questions.[1] (4) Does the application of this technique alter the sample composition (i.e., do the remaining participants in different conditions of loading time differ in sex, age, and education)?

If increasing the loading time of the first questionnaire page works as desired, we should find the first three hypotheses confirmed:

Hypothesis 1: The longer the loading time, the lower the likelihood of people reaching the first page of the study.

Hypothesis 2: The longer the loading time, the smaller the dropout rate.

Hypothesis 3: The longer the loading time, the higher the data quality produced by the remaining sample. In detail, it is expected that the longer the loading time, (3a) the lower the nonresponse to closed-ended items, (3b) the more likely it is that open-ended questions will be answered, and (3c) the longer the answers to the open-ended questions.[2]

In addition, we tested Hypothesis 4 for exploratory reasons:

Hypothesis 4: The loading time influences the composition of the responding sample in terms of sex, age, and education.

STUDY 1

Method
Procedure. When determining the range of manipulating the loading time of the first page of the study, we heeded the recommendations by Shneiderman (1998) and Nielsen (2000), who advised against loading times longer than 10–15 sec and 10 sec, respectively. If the loading time exceeds 10 sec, even patient participants are likely to think that a technical error has occurred (Lazar & Norcio, 1999), or the participants could feel lost (Sears, Jacko, & Dubach, 2000). Consequently, the maximum loading delay was chosen to be 9 sec. In total, there were four levels of delay: (1) no delay, (2) a 3-sec delay, (3) a 6-sec delay, and (4) a 9-sec delay. Adding a physically necessary loading time estimated at 1 sec to the artificially introduced delay, the following loading times resulted in the four experimental conditions: 1, 4, 7, and 10 sec.

The invitation e-mail contained a hyperlink to the first page of the study. The study was about the image of different scientific disciplines. Upon clicking this hyperlink—before the first page of the study appeared—people were randomly assigned to one of four conditions: They needed to wait 1, 4, 7, or 10 sec before the first page of the study was displayed on the screen. The first page stated the aims of the study. On page 2, respondents were asked whether they were taking part in this study for the first time and whether their participation could be considered serious (seriousness check; see Reips, 2002b). On page 3, respondents were asked demographic questions. On page 4, participants were asked an open-ended question on what they consider to be the objective(s) of psychology as a scientific discipline. On pages 5, 7, 9, and 11, participants' attitudes toward the scientific disciplines of psychology, sociology, business studies, and mathematics were measured with semantic differentials, each consisting of 14 pairs of attributes on six-point scales. On pages 6, 8, 10, and 12, the participants indicated their global attitudes toward psychology, sociology, business studies, and mathematics on a feeling thermometer ranging from 0° to 100°. On the final page, they could leave a comment on the study.

Participants. In January 2005, 6,545 students of the University of Vienna were successfully e-mailed an invitation to the study. Their e-mail addresses were gleaned from the online address book of the university. Of those, 1,500 students reached the first page of the questionnaire. Of those, 41 data sets were discarded because they failed the seriousness check (self-reported multiple or insincere participation). Thus, with 1,459 valid data sets, the response rate was 22.3%.

In terms of self-reported demographics, the participating students—both graduate and undergraduate—were on average 26.3 years old ($SD = 6.2$). Within the sample, 53% of the participants were women, 98% were Austrian, and 24% held a college degree.

Results

Likelihood of responding to the study. Table 1 lists the number of people who successfully reached the first page of the study, broken down by loading time of the first page of the study. Eyeballing the data reveals that indeed, more people reached the first page of the study if there was no delay than if there was an artificial delay. In a log-linear approach using the program LEM (Vermunt, 1997a), we examined which model could best explain this pattern of response. To evaluate different models, we used log-likelihood ratio tests and the Bayesian information criterion (BIC) (Vermunt, 1997b). The BIC takes into account explained variance as well as parsimony of the model. The lower the BIC, the better a model fits the data.

Two plausible models that postulate a decline in the number of respondents as the loading time increases were tested against the model that reflects the null hypothesis—namely, that the same number of people reach the first page, regardless of its loading time ($L^2 = 1.82$, $df = 2$, $p = .40$; BIC = 4,052.49). The first alternative model posits a linear decline ($L^2 = 0.73$, $df = 2$, $p = .69$; BIC = 4,051.41) of the number of respondents as loading time increases. The size of the linearly declining effect amounts to an odds ratio (OR) of 0.98. A result of $0 < OR < 1$ denotes a negative effect; OR = 1 denotes a null effect; and $1 < OR < \infty$ denotes a positive effect. Readers who are more familiar with effect size measures such as Cohen's d or Pearson's correlation coefficient r might use the formulae in Sánchez-Meca, Marín-Martínez, and Chacón-Moscoso (2003) to convert OR into their effect size measure of choice. As a rule of thumb, ORs of >1.4 or <0.7 denote a small effect, ORs of >2.5 or <0.4 a medium effect, and ORs of >4.0 or <0.25 a large effect. In the linear case at hand, an OR of 0.98 implies that the expected number of people who call up a study changes by a factor of 0.98 if the loading delay of the first page increases by 1 sec. The second alternative model posits a logarithmic decline of the number of respondents as loading time increases ($L^2 = 0.29$, $df = 2$, $p = .87$; BIC = 4,050.97; OR = 0.96). In a logarithmic model, on the basis of the OR one can calculate the change in the expected number of people who call up a study as the loading delay of the first page increases according to Formula 8.4 in Agresti

(2002, p. 317). Of the three models under consideration, the logarithmic model fit the data best.

Dropout. Table 1 lists the percentages of respondents who dropped out after the first page of the study, broken down by experimental condition. Two plausible models that postulate a decline in dropout as the loading time increases were tested against the model that reflects the null hypothesis—namely, that the same percentages of respondents abandoned the study prematurely, regardless of the loading time of the first page ($L^2 = 3.15$, $df = 3$, $p = .37$; BIC = 5,646.73). The first alternative model posits a linear decline ($L^2 = 2.29$, $df = 2$, $p = .32$; BIC = 5,653.15; OR = 0.98). The OR of 0.98 implies that the odds of a person dropping out prematurely rather than staying until the end of the questionnaire changes by a factor of 0.98 if the loading delay of the first page increases by 1 sec. The second alternative model posits a logarithmic decline ($L^2 = 2.70$, $df = 2$, $p = .26$; BIC = 5,653.57; OR = 0.91) of the percentage of respondents who drop out as loading time increases. The model that reflects the null hypothesis fit the data best.

Answering of closed-ended questions. The study contained 68 closed-ended questions. A linear regression analysis, including only respondents who did not drop out of the study [$\beta = -.04$, $t(1112) = -1.29$, $p = .20$] yielded no significant effect of the loading time on the number of closed-ended questions that were answered (see Table 1).

Answering of open-ended questions. There were two open-ended questions in the study. We examined whether participants were more likely with increasing loading time to answer each of these open-ended questions, and whether any of the people who answered these questions typed in more characters than others (see Table 1). Using a log-linear approach and including only nondropouts, two models that postulate a higher probability of answering the first open-ended question as the loading time increases were tested against the model that reflects the null hypothesis—namely, that the same percentages of people answered this open-ended question ($L^2 = 3.48$, $df = 3$, $p = .32$; BIC = 4,264.50). The first alternative model posits a linear increase ($L^2 = 2.87$, $df = 2$, $p = .24$; BIC = 4,270.90; OR = 1.02) and the second alternative model posits a logarithmic increase ($L^2 = 2.63$, $df = 2$, $p = .27$; BIC = 4,270.66; OR = 1.16) of the percentage of people who answer this

Table 1
Study 1: People Reaching the First Page of the Study, Dropout Rate,
Answering of 64 Closed-Ended Questions, and Answering of 2 Open-Ended
Questions, As Well As Sex, Age (in Years), and Education of Respondents

	Loading Time			
	1 sec	4 sec	7 sec	10 sec
Respondents	387	356	360	356
Dropout	90 (23.3%)	95 (26.7%)	85 (23.6%)	75 (21.1%)
No. of items answered (SD)	67.3 (3.4)	66.7 (7.3)	66.7 (6.7)	66.6 (6.7)
People answering Question 1	234 (78.8%)	193 (73.9%)	218 (79.3%)	225 (80.1%)
Length of Answer 1 (SD)	149.9 (122.1)	135.0 (172.1)	118.0 (79.5)	143.1 (183.3)
People answering Question 2	62 (21.0%)	48 (18.5%)	69 (25.3%)	68 (24.5%)
Length of Answer 2 (SD)	159.9 (208.0)	162.6 (179.1)	145.6 (166.6)	144.6 (138.6)
Female participants	190 (54.3%)	166 (53.0%)	152 (49.7%)	175 (55.7%)
Age of participants (SD)	26.6 (6.1)	26.4 (6.2)	26.1 (6.3)	26.0 (6.0)
Holders of a college degree	96 (27.4%)	89 (28.3%)	51 (16.7%)	68 (21.7%)

open-ended question as loading time increases. Again, the model that reflects the null hypothesis fit the data best.

The same analysis was carried out on the second open-ended question. The model that reflects the null hypothesis ($L^2 = 4.36$, $df = 3$, $p = .23$; BIC = 4,207.57) fit better than the linear ($L^2 = 2.62$, $df = 2$, $p = .27$; BIC = 4,212.82; OR = 1.03) and the logarithmic ($L^2 = 2.59$, $df = 2$, $p = .27$; BIC = 4,212.79; OR = 1.24) models.

Linear regression analyses including only respondents who did not drop out of the study and who answered the respective open-ended question (see Table 1) yielded no significant effect of the loading time on the number of characters that were entered in answer to the first [$\beta = -.06$, $t(868) = -1.80$, $p = .07$] and the second [$\beta = -.05$, $t(245) = -0.70$, $p = .48$] closed-ended questions.

Sample composition. With regard to the percentage of women among the respondents (see Table 1), a model that posits no difference among the four loading-time conditions fit the data best ($L^2 = 2.51$, $df = 3$, $p = .47$; BIC = 5,334.02). A linear regression analysis yielded no effect of the loading time on the age of the respondents [$\beta = -.04$, $t(1284) = -1.43$, $p = .15$]. With regard to the percentage of holders of a college degree among the respondents, the model that posits no difference among the four loading time conditions did not fit the data ($L^2 = 15.89$, $df = 3$, $p = .001$; BIC = 4,972.24). Exploratory analyses showed that a model that posits a linear decline of the percentage of college degree holders as loading time increases ($L^2 = 8.47$, $df = 2$, $p = .02$; BIC = 4,971.98; OR = 0.95) fit the data slightly better, but in absolute terms it did not provide a significant fit, either.

STUDY 2

Method

Procedure. Study 2 used the same procedure and materials as in Study 1. However, this time members of a university-based online panel were invited to participate. An online panel is a pool of people who have agreed to occasionally take part in Web-based studies (see Göritz, 2007). The panelists (i.e., members of the online panel) were volunteers from all walks of life.

Participants. In January 2005, 1,666 panelists were successfully e-mailed an invitation to the study. Of those, 878 people reached the first page of the questionnaire. Of those, 35 data sets were discarded because they failed the seriousness check (self-reported multiple or

insincere participation). Thus, with 843 valid data sets, the response rate was 50.6%. The responding panelists were on average 33.3 years old ($SD = 10.9$). Within the sample, 46% of the participants were women, 94% were German, and 39% held a college degree.

Results

Likelihood of responding to the study. Table 2 lists the number of people who successfully reached the first page of the study, broken down by loading time of the first page of the study. The model that reflects the null hypothesis—namely, that the same number of people reach the first page regardless of its loading time—does not fit the data ($L^2 = 7.60$, $df = 2$, $p = .02$; BIC = 2,996.98). The first alternative model posits a linear decline ($L^2 = 1.29$, $df = 2$, $p = .52$; BIC = 2,337.72; OR = 0.97) and the second a logarithmic decline ($L^2 = 0.12$, $df = 2$, $p = .94$; BIC = 2,336.55; OR = 0.90) of the number of people who reached the first page as loading time increased. Of the two alternative models, the logarithmic model fit the data best, but in terms of statistical significance it was not superior to the linear model.

Dropout. Table 2 lists the percentages of respondents who dropped out after the first page of the study, broken down by experimental condition. The model that reflects the null hypothesis—namely, that the same percentage of respondents abandon the study prematurely, regardless of the loading time of the first page ($L^2 = 2.24$, $df = 3$, $p = .52$; BIC = 2,996.98)—fit the data better than did either the model of a linear decline in dropout rate as loading time increased ($L^2 = 2.16$, $df = 2$, $p = .34$; BIC = 3,003.63; OR = 0.99) or the model of a logarithmic decline ($L^2 = 2.15$, $df = 2$, $p = .34$; BIC = 3,003.62; OR = 0.93).

Answering of closed-ended questions. A linear regression analysis including only nondropouts [$\beta = -.06$, $t(729) = -1.49$, $p = .14$] yielded no significant effect of the loading time on the number of answered closed-ended questions (see Table 2).

Answering of open-ended questions. Using a log-linear approach and including only nondropouts (see Table 2), two models that postulate a higher probability of answering the first open-ended question as the loading time increases were tested against the model that reflects the null hypothesis—namely, that the same percentage of people answer this open-ended question ($L^2 = 7.12$,

Table 2
Study 2: People Reaching the First Page of the Study, Dropout Rate, Answering of 64 Closed-Ended Questions, and Answering of 2 Open-Ended Questions, As Well As Sex, Age (in Years), and Education of Respondents

	Loading Time			
	1 sec	4 sec	7 sec	10 sec
Respondents	245	206	199	193
Dropout	35 (14.3%)	23 (11.2%)	31 (15.6%)	23 (11.9%)
No. of items answered (SD)	67.6 (0.7)	67.4 (1.3)	67.5 (0.7)	67.4 (1.3)
People answering Question 1	182 (86.7%)	149 (81.4%)	153 (91.1%)	145 (85.3%)
Length of Answer 1 (SD)	125.9 (102.0)	113.2 (112.6)	133.2 (119.3)	179.9 (191.6)
People answering Question 2	30 (14.4%)	26 (14.2%)	25 (14.9%)	29 (17.1%)
Length of Answer 2 (SD)	145.5 (210.4)	122.8 (148.9)	100.3 (88.6)	112.7 (97.1)
Female participants	106 (45.9%)	87 (43.9%)	94 (49.5%)	81 (44.0%)
Age of participants (SD)	33.4 (11.2)	32.9 (10.7)	32.9 (11.1)	33.8 (10.8)
Holders of a college degree	91 (40.1%)	80 (40.4%)	59 (31.4%)	78 (42.9%)

$df = 3$, $p = .07$; BIC $= 2,618.17$). The first alternative model posits a linear increase ($L^2 = 6.94$, $df = 2$, $p = .03$; BIC $= 2,624.57$; OR $= 1.01$) and the second a logarithmic increase ($L^2 = 7.00$, $df = 2$, $p = .03$; BIC $= 2,624.64$; OR $= 1.09$) of the percentage of people who answered this open-ended question as loading time increased. Only the model that reflects the null hypothesis fit the data. The same analysis was carried out on the second open-ended question. Again, the model that reflects the null hypothesis ($L^2 = 0.71$, $df = 3$, $p = .87$; BIC $= 2,643.73$) fit the data better than the linear ($L^2 = 0.18$, $df = 2$, $p = .91$; BIC $= 2,649.80$; OR $= 1.02$) or the logarithmic ($L^2 = 0.09$, $df = 2$, $p = .96$; BIC $= 2,649.71$; OR $= 1.20$) model.

Linear regression analyses including only respondents who did not drop out of the study and who answered the respective open-ended question (see Table 2) yielded no effect of the loading time on the number of characters that were entered in answer to the first [$\beta < .01$, $t(627) = 0.05$, $p = .96$] or the second [$\beta = -.10$, $t(108) = -1.04$, $p = .30$] closed-ended question.

Sample composition. With regard to the percentage of women among the respondents (see Table 2), the model that posits no difference among the four loading time conditions fit the data best ($L^2 = 1.54$, $df = 3$, $p = .67$; BIC $= 3,334.27$). A linear regression analysis yielded no effect of the loading time on the age of the respondents [$\beta = .01$, $t(803) = 0.28$, $p = .78$]. With regard to the percentage of holders of a college degree among the respondents, the model that posits no difference among the four loading time conditions fit the data best ($L^2 = 6.10$, $df = 3$, $p = .11$; BIC $= 3,266.45$).

DISCUSSION

Two experiments examined the usefulness of artificially delaying the loading time of the first page in a Web-based study, which is a variant of the high-hurdle technique (see Reips, 2002b). The idea pursued in this technique was to screen out respondents with a low level of motivation at the very beginning of a study. Participants who remained in the study despite this hurdle were expected to produce data of higher quality because of their high level of motivation. The increase in data quality should be reflected in a lower likelihood of prematurely abandoning the study, lower nonresponse to both closed- and open-ended questions, and longer answers to open-ended questions.

Hypothesis 1, which stated that the longer the loading time, the smaller the likelihood of invited people responding to the study, was supported in both experiments. Hypothesis 2, which postulated that the longer the loading time, the smaller the rate of dropout, was not supported in either experiment. Moreover, data quality in the form of response behavior to closed-ended and open-ended questions proved to be independent of the loading time of the first page of the study. Hence, Hypotheses 3a, 3b, and 3c were not supported in either experiment. Finally, we explored whether the loading time influenced the composition of the responding sample. No such influences on sex, age, or education were found. Thus, Hypothesis 4 was rejected. The inconsequentially small size of the effects postulated in Hypotheses 2–4, combined with the relatively large sam-

ples in the two studies, shows that the null hypotheses were confirmed not simply because of low power, but because of the actual insubstantiality of the alternative effects.

The findings warrant the conclusion that increasing the loading time—as one variant of the high-hurdle technique—does not fulfill its promise. High loading times lower people's likelihood of responding to a study, but at the same time they do not enhance the quality of the data. A high loading time constitutes extra respondent burden, but not a burden that is suitable for sifting respondents with higher motivation from those with lower motivation. Instead, high loading times alienate potential respondents, regardless of their level of motivation. A likely mechanism is that respondents, independent of their level of motivation, take a long loading time as indicative of a technical problem with the study. Although from testing Hypotheses 2–4 there was no hint that an increased loading time leads to selection bias, we might still be at risk of losing particular clienteles of respondents more than others. Such a hidden bias might be detected if dependent variables other than the ones examined in this study were investigated.

We believe that examining this variant of the high-hurdle technique in two experiments using big samples from different backgrounds is a unique contribution of this study. The remarkable consistency of results obtained in these two studies speaks for the generalizability of the findings. However, our work is not without limitations. Both samples' education was above average, and both studies were restricted; that is, they relied on an invited sample and were not freely accessible by any Web user who happened to come across the Web site. Studies need to be conducted to find out whether artificially delaying the loading of pages works better with unrestricted (i.e., uninvited) samples.

Moreover, in both experiments, the maximum loading time was 10 sec. Perhaps with a loading time longer than 10 sec, desired effects on the data quality will materialize after all. Furthermore, in future studies, indicators of data quality other than the ones examined in the studies at hand might be used, such as the honesty with which questions were answered. Finally, Reips (2002b) suggested that loading not only of the first page of the study should be artificially delayed, but loading of the first *few* pages in decreasing fashion. If the proposed mechanism of respondents attributing the long waiting time to technical flaws applies, delaying loading of the first few pages should not yield better results than delaying loading of the first page only. Moreover, when only delaying the loading of the first page—as was done in the two studies at hand—some of the desired effects should already be discernible. However, because this proved not to be the case, delaying the loadings of the first few pages likely would not live up to its promise, either. It remains for future studies to resolve this question empirically.

To sum up, at least with restricted samples, artificially delaying the loading of the first page of a study is counterproductive and hence not recommended. Such a delay does not yield any gain in terms of data quality, but instead incurs loss: It decreases response to the study, requires that programmers spend extra time to implement the loading delay, and may scare away volunteers from participating in future studies run by the same research organization,

because these volunteers suspect incompetence in technical matters.

REFERENCES

AGRESTI, A. (2002). *Categorical data analysis.* New York: Wiley.

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

CRAWFORD, S. D., COUPER, M. P., & LAMIAS, M. J. (2001). Web surveys: Perceptions of burden. *Social Science Computer Review*, **19**, 146-162.

GÖRITZ, A. S. (2004). The impact of material incentives on response quantity, response quality, sample composition, survey outcome, and cost in online access panels. *International Journal of Market Research*, **46**, 327-345.

GÖRITZ, A. S. (2005). Contingent versus unconditional incentives in WWW-studies. *Metodološki Zvezki—Advances in Methodology & Statistics*, **2**, 1-14.

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

GÖRITZ, A. S. (2007). Using online panels in psychological research. In A. N. Joinson, K. Y. A. McKenna, T. Postmes, & U.-D. Reips (Eds.), *The Oxford handbook of Internet psychology* (pp. 473-485). Oxford: Oxford University Press.

GÖRITZ, A. S., & SCHUMACHER, J. (2000). The WWW as a research medium: An illustrative survey on paranormal belief. *Perceptual & Motor Skills*, **90**, 1195-1206.

HEERWEGH, D., & LOOSVELDT, G. (2002). Web surveys: The effect of controlling survey access using PIN numbers. *Social Science Computer Review*, **20**, 10-21.

HEERWEGH, D., & LOOSVELDT, G. (2003). An evaluation of the semiautomatic login procedure to control Web survey access. *Social Science Computer Review*, **21**, 223-234.

LAZAR, J., & NORCIO, A. (1999). To err or not to err, that is the question: Novice user perception of errors while surfing the Web. In M. Khosrowpouv (Ed.), *Proceedings of the Information Resource Management Association 1999 International Conference* (pp. 321-325). Hershey, PA: Idea Group.

NIELSEN, J. (2000). *Designing Web usability: The practice of simplicity.* Indianapolis: New Riders.

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**, 217-226.

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). San Diego: Academic Press.

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

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

REIPS, U.-D. (2002c). Theory and techniques of conducting Web experiments. In B. Batinic, U.-D. Reips, & M. Bosnjak (Eds.), *Online social sciences* (pp. 229-250). Seattle: Hogrefe & Huber.

SÁNCHEZ-MECA, J., MARÍN-MARTÍNEZ, F., & CHACÓN-MOSCOSO, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods*, **8**, 448-467.

SCHMIDT, W. C. (2000). The server-side of psychology Web experiments. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 285-310). San Diego: Academic Press.

SCHWARZ, S., & REIPS, U.-D. (2001). CGI versus JavaScript: A Web experiment on the reversed hindsight bias. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (pp. 75-90). Lengerich, Germany: Pabst Science.

SEARS, A., JACKO, J., & DUBACH, E. (2000). International aspects of WWW usability and the role of high-end graphical enhancements. *International Journal of Human-Computer Interaction*, **12**, 243-263.

SHNEIDERMAN, B. (1998). *Designing the user interface: Strategies for effective human-computer interaction* (3rd ed.). Reading, MA: Addison-Wesley.

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

STIEGER, S., REIPS, U.-D., & VORACEK, M. (2007). Forced-response in online surveys: Bias from reactance and an increase in sex-specific dropout. *Journal of the American Society for Information Science & Technology*, **58**, 1653-1660.

VERMUNT, J. K. (1997a). *LEM: A general program for the analysis of categorical data* [Computer program]. Tilburg, The Netherlands: Tilburg University.

VERMUNT, J. K. (1997b). *Log-linear models for event histories.* Thousand Oaks, CA: Sage.

WENZEL, O. (2001). *Webdesign, Informationssuche und Flow: Nutzerverhalten auf unterschiedlich strukturierten Websites* [Web design, search for information, and flow: User behavior on differently structured Web sites]. Lohmar, Germany: Eul.

NOTES

1. Of course, a skipped item or a briefer answer does not imply lower data quality in every single case—the opposite might even be true. However, on average, the more complete a filled-out questionnaire and the longer an answer to an open-ended question, the greater the likelihood a respondent took care when answering.

2. For two reasons, we rejected the viewing time of the questionnaire as an indicator of data quality. The first reason is not specific to this study. There are people with a low level of motivation at both ends of the spectrum of viewing times—that is, click-throughs (i.e., people answering items in a short time without reading questions carefully) and people who interrupt their participation. Thus, without cleaning out click-throughs—which always involves somewhat arbitrary decisions—a longer viewing time does not imply higher data quality. The second reason is specific to this study. With the variation of loading time of the first page, viewing time as in indicator of data quality has low internal validity, because an alternative account is possible: If the first page loads slowly, participants' viewing time of the next page(s) might be increased because they have occupied themselves with other things while waiting for the first page to appear. Because these people are less likely to look attentively at the screen, they return to the study later than those who did not find the time to occupy themselves with other things while waiting. This difference in viewing time could have carried over onto the next pages of the study, because people who waited a long time for the first page to load might have assumed that the following pages would also load slowly, and thus they might have turned to an occupation on the side.