

A comparison of World-Wide Web and paper-and-pencil personality questionnaires

FRANCES ANNIE PETTIT
York University, Toronto, Ontario, Canada

Does the manifestation of response set effects in World-Wide Web (WWW) questionnaire responses differ from that in paper-and-pencil (PP) questionnaire responses? Two thousand six hundred and forty-nine volunteers responded to a WWW questionnaire, and 458 volunteers responded to a PP questionnaire consisting of five personality scales. Five response sets were examined. For four of them—random response, item nonresponse, extreme response, and acquiescent response—no statistically significant differences between PP and WWW data were found. The PP administration elicited a statistically higher number of errors (uncodable responses), although η^2 was less than .02. An analysis of interitem and interscale correlations did not differentiate WWW and PP data. Data from the Marlowe–Crowne Social Desirability Scale, the Perfectionist Self-Presentation Scale, and the Computer Anxiety Scale did not differentiate WWW and PP data in terms of either means or scale reliabilities. It was concluded that WWW data may be comparable to PP data and that the WWW is a potentially useful and valid data collection tool.

The World-Wide Web (WWW) is a network of interlinked computer servers that is increasingly widely accessible and includes users from all ages, ethnic origins, and geographic locations. Despite the advantages of the e-mail and computer-mediated administration methods, such as the reduced number of data entry errors and a wider range of subjects, paper and pencil (PP) has continued to remain the most popular questionnaire administration method. However, data collection via the WWW method incorporates many advantages from both the computer-mediated and the e-mail procedures, as well as savings of time. There are disadvantages, such as the lack of control, but the advantages make this method very enticing.

At least as early as 1995, researchers began to seriously test the feasibility of using the WWW, with moderate success in the speed of data collection (Welch & Krantz, 1996). More current attempts have illustrated that personality questionnaire data via the WWW can occur extremely rapidly (Pettit, 1999). However, in these studies, the feasibility of WWW data collection was examined, rather than the quality of the data collected via the WWW.

One way to evaluate the quality of data is to examine response sets. In questionnaires, response sets are patterns of responding that are independent of the questions. Cronbach (1946) indicated that “because response sets permit persons with the same knowledge or attitude or ability to receive different scores, response sets always lower the logical validity of a test” (p. 484). Five common response effects, including random responding, response errors, item nonresponse, acquiescent responding, and extreme responding, will be quickly reviewed here.

Random responding occurs when participants haphazardly provide responses without reading the question. Using students to compare computer-mediated and PP administrations, Beach (1989) found that random response rates were higher for a computer-mediated than for a PP-administered attitude questionnaire. Only 4% of the PP participants exhibited symptoms of random responding, as compared with 10% of the computer-mediated participants. As measured by the F scale of the MMPI, Locke and Gilbert (1995) determined that random response rates were higher in both computer-mediated and PP administration groups than in interview groups.

Response errors occur when a response to an item has been generated but the response cannot be used (e.g., illegible handwriting). In a questionnaire about health and personal characteristics, Kiesler and Sproull (1986) found that the e-mail administration (0%) produced a lower error rate than did the PP administration (5.3%) because it was impossible for e-mail participants to provide illegible responses.

Item nonresponse is one of the most commonly studied response sets. In a questionnaire about the effectiveness of a computer-training session, Webster and Compeau (1996) observed a higher average rate of missing responses in the computer-mediated administration than in the PP administration. However, Locke and Gilbert (1995) did not observe any difference in the rate of incomplete questionnaires when comparing PP and computer-mediated methods.

Acquiescence is characterized by responses that show agreement with a question regardless of the question's content. This response set is characterized by responses with an unusually higher number of *true* and *strongly agree* responses. Kiesler and Sproull (1986) observed no significant differences between the PP and the e-mail data in a health and personal characteristics questionnaire.

Correspondence should be addressed to F. A. Pettit, 50 Stately Way, Thornhill, ON, L3T 3Z8 Canada (e-mail: pettitconsulting@rogers.com).

Finally, extreme responding occurs when participants favor the anchors of a scale, such as the *strongly agree* or *strongly disagree* options on a Likert scale. Sproull (1986) found that 72% of the e-mail-administered questions elicited a higher extreme response rate than did the PP administration, using a sample of business people. On the other hand, Helgeson and Ursic (1989) studied decision processes in undergraduate students and found no significant differences in extreme responses between the PP and the computer-mediated administrations.

Social desirability is another pattern of responding that might differ between WWW and PP data. Two studies using the Marlowe-Crowne Social Desirability Scale (MCSDS) found that social desirability was greater in a PP administration than in a computerized administration. Martin and Nagao (1989) studied students participating in a job application study, and Kiesler and Sproull (1986) studied university members participating in a health study. It is not clear why social desirability was less in the e-mail administration, but one hypothesis is the increased social distance provided by e-mail.

How do responding populations differ between the WWW and the general population? In 1997, Schmidt recognized that WWW users are not typical of the general population in that they tend to be younger and male and to have a higher average education and socioeconomic status. Also in 1997, Smith and Leigh found that their WWW sample did not differ from their student sample in terms of marital status, ethnicity, age, or education, although the majority of the WWW volunteers were men.

Even though the demographics of WWW respondents can differ from those of PP respondents, it is more important to examine whether the research results differ. Krantz, Ballard, and Scher (1997) compared WWW volunteers with psychology students and found no differences. However, the topic (female attractiveness) was limited, and the PP and WWW sources were different. Pasveer and Ellard (1998) examined questionnaire data about self-trust obtained via the WWW and PP. Using a sample of over 1,400 individuals in the WWW group and over 700 students in the PP group, the researchers demonstrated that the psychometric properties of the scale were nearly identical. Again, however, the topic of interest was limited, and the samples were from different sources.

A number of issues need to be addressed before it can be concluded that the quality of WWW data is similar to the quality of PP data. First and most important, the quality of WWW data has not yet been examined. Second, generalizability studies must be carried out in a wide range of areas to ensure that generalizability is not limited to specific areas. Third, psychological aspects that are directly applicable to research psychologists must be studied. Fourth, samples must be obtained from the same source. Finally, research must be carried out with nonstudent samples.

The purpose of the present study was to discover whether WWW data can adequately replicate PP data. Generalizability was evaluated in terms of the five response set ef-

fects discussed, as well as means, correlations, and the psychometric properties of scales. Furthermore, the research included respondents from the same source, nonstudent respondents, and questionnaires with psychological content.

METHOD

Participants

The WWW sample consisted of 2,649 self-selected adults worldwide who completed the WWW questionnaire. A subsample of 190 individuals volunteered their mailing address first (in order to prepare for a possible confound introduced by assigning WWW participants to the PP administration only if they provided an address).

The WWW volunteers consisted of 1,645 women and 905 men. The average age was 30 years ($SD = 11.4$). Almost 77% of the respondents reported having at least attended college. Approximately 35% were students.

The PP sample consisted of 459 adults making up three groups. The first group consisted of 184 adults residing across the United States and Canada. These volunteers were recruited in exactly the same manner as were the respondents in the WWW sample, except that a PP questionnaire was mailed to each of these volunteers. A second group of 101 nonstudent adults was recruited from a variety of places of employment. Finally, a third group of 164 full-time and part-time undergraduate university students was recruited from within York University, Toronto, Canada.

The PP sample consisted of 274 women and 165 men. The average age was 29 years ($SD = 10.5$). More than 81% of the respondents reported having at least attended college. Approximately 48% were students.

No incentives were offered.

WWW Recruitment and Administration Procedures

Several procedures were employed in order to attract volunteers to the WWW questionnaire. The questionnaire was registered on numerous search engines (e.g., Yahoo, Excite, Infoseek), on-line research lists (e.g., the American Psychological Society), and newsgroups, and banners were also used.

Upon reaching the Web site, potential participants were greeted by an introductory screen that indicated that the researcher was interested in examining differences in people's answers to WWW and PP questionnaires. The participants were guaranteed anonymity, were told that they did not have to answer every question, and were told that they could withdraw at any time. My name, affiliation, and e-mail address were included here, and several participants took advantage of this.

Those participants who did not wish to participate left the site as they would any Web site in which they were not interested. The participants who wished to view the questionnaire clicked on the provided link. The questionnaire indicated that they were to use the scales provided to answer the questions quickly. If the participants chose to submit their responses, they clicked on another link that automatically recorded their responses in a database and took the participants to a debriefing and thank you page. This page thanked them for their participation, further described the purpose of the experiment, and gave my e-mail address again. Additional participants chose to e-mail me at this point.

After all the WWW respondents had been obtained, the welcome page was altered so that the respondents were requested to e-mail the researcher with their postal addresses prior to participating. Canadian and American respondents who did so were then randomly assigned to respond to either the WWW or the PP administration. The participants either were mailed a questionnaire or received one e-mail indicating the address of the WWW questionnaire. Those participants

who were mailed a questionnaire received one e-mail indicating that a questionnaire had been sent to the address they indicated and two reminder e-mails. The participants were not required to indicate their e-mail addresses on the questionnaire they submitted. However, nearly every participant did submit his or her address and was sent the debriefing material. My e-mail and physical addresses were contained in the questionnaire material.

Both the PP and the WWW questionnaires could be completed within approximately 20 min. The participants in both the PP and the WWW samples were aware that they could backtrack to previous items, change responses to previous items, see how many items had been completed and were left to be completed, skip items, and respond at their own pace. The WWW questionnaire allowed the participants to leave any item unanswered. The participants were not forced to complete every item, nor did they have responses imputed prior to answering items. The PP questionnaire was a high-quality print-out of the WWW questionnaire.

Scales and Measures

Computer Attitude Scale (CAS; Loyd & Gressard, 1984). The CAS consists of ten 6-point Likert scaled items, half of which are reverse coded. High scores indicate low anxiety. Cronbach's internal consistency coefficient is $\alpha = .86$. This scale was chosen because the items most closely reflected the anxiety construct under examination and it had been tested with adults.

Perfectionist Self-Presentation Scale (PSPS; Hewitt & Flett, 1995). The PSPS measures the fear of making mistakes in front of other people and the reluctance to disclose things in public. The 27 items are measured on a 7-point Likert scale, 5 items of which are reverse keyed. High scores indicate high self-presentation perfectionism. This scale was selected because it would indicate whether WWW participants are more open than PP participants.

Marlowe-Crowne Social Desirability Scale (MCSDS; Marlowe & Crowne, 1960). The MCSDS consists of 33 true/false items. High scores indicate high degrees of social desirability. As reported by the authors, internal consistency is $\alpha = .88$, and test-retest reliability is $.89$. This scale was selected because it would provide a comparison measure to past research (Kiesler & Sproull, 1986; Martin & Nagao, 1989).

Pettit WWW Random Responding Scale (PWRRS; Pettit, 1998). The PWRRS is a 10-item true-false measure of random responding in which half of the items are reverse keyed. After recoding the negatively keyed items, the number of *true*s is summed and multiplied by 100 to obtain a total score. Scores of 30 or greater indicate that the participant is likely responding randomly. The scale was developed and validated using 771 WWW participants. The validation sample comprised 64% women and 36% men, the average age was

31 years, and 34% were students. The mean score was $M = 20.1$ ($SD = 52.0$). This scale was developed because permission could not be obtained to use common scales on the WWW.

The response effect rates were operationalized as follows.

Random response: score on the PWRRS.

Response errors: the total number of items for which (1) two or more options were selected, (2) it was not obvious which option was finally selected, (3) the response was illegible, (4) the response was inappropriate or impossible, or (5) a demographic combination was unlikely.

Nonresponse: the total number of items for which no response was provided.

Extreme response: the percentage of answered items for which either *strongly agree* or *strongly disagree* was selected.

Acquiescent response: Percentage of answered items for which the response selected was *strongly agree*, *slightly agree*, *agree*, or *true*.

RESULTS

Prior to carrying out this research, it was determined that a sample size of 400 was required to obtain satisfactory power. Because large sample sizes, such as were used here, lead to statistical significance even with minute effect sizes, effect sizes were only reported if $p < .01$. Power levels have not been provided, because all of the nonsignificant tests were associated with extremely minute effect sizes, which would not have been significant even with a sample size twice as large.

Random Sets

Table 1 indicates that no significant differences were found between the WWW and the PP data for mean random responding, extreme responding, acquiescent responding, or nonresponse (logarithmically transformed owing to skewness). There was no interaction between the method of administration and providing an address for any of these variables, so no obvious confound was introduced by requesting addresses from PP and WWW participants. Data from the 51 random responders were eliminated from further analyses. Based on the criterion of more than three errors, a further 10 questionnaires were eliminated.

The mean number of WWW errors was greater than the mean number of PP errors (log of error was used owing to skewness; see Table 1). The interaction between providing

Table 1
Statistics for Response Sets

Variable	WWW		PP		F
	M	SD	M	SD	
Random	1.90	6.60	2.60	7.60	$F(1,3079) = 2.09, p = .15$ $F(1,3079) = 5.77, p = .02^*$
Errors	0.14	0.37	0.67	0.97	$F(1,2919) = 20.37, p < .001, r^2 = .01$ $F(1,2919) = 5.55, p = .004, r^2 = .01^*$
Nonresponse	0.22	1.08	1.10	2.03	$F(1,2644) = 0.05, p = .83$ $F(1,2644) = 3.82, p = .05^*$
Extreme	35.0	20.70	29.20	19.70	$F(1,2639) = 0.54, p = .46$ $F(1,2639) = 3.30, p = .04^*$
Acquiescence	48.3	8.05	49.50	8.40	$F(1,2652) = 4.37, p = .04$ $F(1,2652) = 0.36, p = .55^*$

Note—WWW, World-Wide Web; PP, paper and pencil. *Statistics for interaction between method of administration and providing addresses.

Table 2
Error Means and Standard Deviations for Interaction Between
Providing Addresses and Method of Administration

Anonymous	WWW			PP		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Yes	0.14	0.37	2,316	0.54	0.91	238*
No	0.13	0.35	185	0.85	1.02	181

Note—WWW, World-Wide Web; PP, paper and pencil. *These participants were recruited in person. All other participants were recruited via the WWW.

addresses and the method of administration was also statistically significant. As is shown in Table 2, PP volunteers recruited from the WWW produced nearly 60% more errors than did PP volunteers recruited in the traditional manner.

Bivariate Relationships

Because researchers are more commonly interested in studying relationships between variables, it was important to determine whether relationships within the WWW data were consistent with the relationships within the PP data. A discriminant analysis based on 28 items from the CAS, the PSPS, and the MCSDS indicated that the WWW covariance matrix differed from the PP covariance matrix [$\chi(406) = 552.94, p < .0001$] and indicated that the WWW covariance matrix differed from the PP covariance matrix.

To examine this difference in more detail, a Z-test examination of each of the 378 interitem correlations was carried out to reveal a nonsignificant mean level of statistical significance ($p = .484, SD < 0.317$). After Bonferroni adjustment, no more than a chance proportion (1.1%) of the 378 Z-tests was statistically significant at $p < .01$. Furthermore, none of the intercorrelations between the scale scores for the MCSDS, the PSPS, and the CAS was statistically significant.

Scale Results

Table 3 indicates that the mean scores on the MCSDS, the PSPS, and the CAS did not differentiate the WWW sample from the PP sample. Measures of internal consistency also did not differentiate the WWW data from the PP data.

DISCUSSION

Can psychology questionnaire data obtained via the WWW generalize to psychology questionnaire data ob-

tained via PP? When looking at random responding, item nonresponse, extreme responding, acquiescent responding, univariate and bivariate results, and scale properties, the answer is yes. The only response set that was different led to the conclusion that WWW administrations produce fewer errors.

The finding of similar random responding rates parallels the research of Locke and Gilbert (1995), who also failed to differentiate computerized administrations from PP administrations. Item nonresponse also failed to differentiate the WWW data from the PP data. But, as with the studies reviewed in the introduction (Webster & Compeau, 1996), this study found that the incomplete questionnaire rate for the PP administration (50%) was twice that of the computerized administration (25%). Past research with other types of computerized administration has also demonstrated that extreme responding (Helgeson & Ursic, 1989) and acquiescent responding (Kiesler & Sproull, 1986) rates in WWW data are the same as those in PP data.

Naturally, researchers are interested in univariate results, but most researchers place a greater emphasis on relationship among variables. It is encouraging that most correlations did not differ between WWW and PP administrations and that the psychometric properties of the scales were similar.

There are some limitations of this research. Even though computers are becoming increasingly common, access to the WWW is still restricted. As computers become even more commonplace, the characteristics of WWW volunteers may more accurately reflect general population characteristics. A further limitation is that the WWW participants were self-selected, a fact that was clear given the greater rate of responding by women even though WWW users tend to be men. It is more important to recognize that the volunteers, self-selected or not, produced WWW and PP data that could not be differentiated. In addition, this study in no way represents the entire range of psychological questionnaires that could differentiate PP data from WWW data.

This research is important for a number of reasons. First, volunteers from the WWW participated in both the PP and the WWW administrations. This is a rarity in the research and a fact that made it possible to attribute differences to the methodology, rather than to the sample source. Second, by comparing WWW volunteers with traditional PP volunteers, it was possible to determine that the data given by the WWW users was not inherently dif-

Table 3
Scale Statistics

Variable	WWW			PP			<i>F</i>
	<i>M</i>	<i>SD</i>	α	<i>M</i>	<i>SD</i>	α	
MCSDS	1.33	0.21	.60	1.36	0.22	.62	$F(1,2652) = 0.42, p = .52$
PSPS	4.00	1.23	.91	3.93	1.09	.88	$F(1,2652) = 0.05, p = .82$
CAS	1.85	0.72	.86	1.98	0.84	.91	$F(1,2652) = 0.01, p = .94$

Note—WWW, World-Wide Web; PP, paper and pencil.; MCSDS, Marlowe-Crowne Social Desirability Scale; PSPS, Perfectionist Self-Presentation Scale; CAS, Computer Attitude Scale.

ferent from the data given by the PP volunteers. Third, although research in this area has been extremely popular in business research and market research, it is long overdue in the field of psychology. Fourth, this study examined univariate and bivariate relationships, scale scores, and internal consistency.

This study has demonstrated that data collection via the WWW can be extremely advantageous. However, because it is technologically impossible to restrict access to any Web site, experiments must be restricted to those that are sensitive to age and maturity. The time required to set up and advertise a Web site could easily take more than 6 months, but the time required to recruit volunteers is very short.

Future Research

With the WWW growing at an exponential rate and the corresponding increase in the potential participant pool, it is easy to access a wider range of people than ever before. In particular, cross-cultural research would benefit greatly. Before any definitive conclusions can be made about the characteristics of response effects, further research must replicate the results of this study in terms of various psychological measures and topics. The few measures in this research were nonintrusive and were not a representative sample of all possible measures. And just because this WWW data generalized to PP data does not mean that all research using the WWW is generalizable to non-WWW procedures. Generalizability studies are still required in many fields.

Conclusion

Carrying out data collection over the WWW means that the potentially high cost of paper and stamps is eliminated, the questionnaire can be aesthetically pleasing, typographic or grammatical errors can be corrected without reprinting the entire set of questionnaires, data entry and the associated errors are completely eliminated, and volunteers are recruited extraordinarily quickly. Clearly, these advantages, combined with the fact that PP data are not qualitatively different from WWW data, indicate that the WWW can be a good place to recruit volunteers and to administer psychological questionnaires for experiments.

REFERENCES

- BEACH, D. A. (1989). Identifying the random responder. *Journal of Psychology*, **123**, 101-103.
- CRONBACH, L. J. (1946). Response sets and test validity. *Educational & Psychological Measurement*, **6**, 475-494.
- HELGESON, J. G., & URSIC, M. L. (1989). The decision process equivalency of electronic versus pencil-and-paper data collection methods. *Social Science Computer Review*, **7**, 296-310.
- HEWITT, P., & FLETT, G. (1995). *Perfectionistic Self-Presentation Scale*. Unpublished test, York University, Toronto.
- KIESLER, S., & SPROULL, L. S. (1986). Response effects in the electronic survey. *Public Opinion Quarterly*, **50**, 402-413.
- KRANTZ, J. H., BALLARD, J., & SCHER, J. (1997). Comparing the results of laboratory and World-Wide Web samples on the determinants of female attractiveness. *Behavior Research Methods, Instruments, & Computers*, **29**, 264-269.
- LOCKE, S. D., & GILBERT, B. O. (1995). Method of psychological assessment, self-disclosure, and experiential differences: A study of computer, questionnaire, and interview assessment formats. *Journal of Social Behaviour & Personality*, **10**, 255-263.
- LOYD, B. H., & GRESSARD, C. (1984). Reliability and factorial validity of computer attitude scales. *Educational & Psychological Measurement*, **44**, 501-505.
- MARLOWE, D., & CROWNE, D. P. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, **24**, 349-354.
- MARTIN, K. L., & NAGAO, D. H. (1989). Some effects of computerized interviewing on job applicant responses. *Journal of Applied Psychology*, **74**, 72-80.
- PASVEER, K. A., & ELLARD, J. H. (1998). The making of a personality inventory: Help from the WWW. *Behavior Research Methods, Instruments, & Computers*, **30**, 309-313.
- PETTIT, F. A. (1998). *Pettit WWW Random Responding Scale*. Unpublished test, York University, North York, ON.
- PETTIT, F. A. (1999). Exploring the use of the World Wide Web as a psychology data collection tool. *Computers in Human Behavior*, **15**, 67-71.
- SCHMIDT, W. C. (1997). World-Wide Web survey research: Benefits, potential problems, and solutions. *Behavior Research Methods, Instruments, & Computers*, **29**, 274-279.
- SMITH, M. A., & LEIGH, G. (1997). Virtual subjects: Using the Internet as an alternative source of subjects and research environment. *Behavior Research Methods, Instruments, & Computers*, **29**, 496-505.
- SPROULL, L. S. (1986). Using electronic mail for data collection in organizational research. *Academy of Management Journal*, **29**, 159-169.
- WEBSTER, J., & COMPEAU, D. (1996). Computer-assisted versus paper-and-pencil administration of questionnaires. *Behavior Research Methods, Instruments, & Computers*, **28**, 567-576.
- WELCH, N., & KRANTZ, J. H. (1996). The World-Wide Web as a medium for psychoacoustical demonstrations and experiments: Experience and results. *Behavior Research Methods, Instruments, & Computers*, **28**, 192-196.