

The making of a personality inventory: Help from the WWW

KAREN A. PASVEER and JOHN H. ELLARD
University of Calgary, Calgary, Alberta, Canada

Data collected electronically from the World-Wide Web (WWW) in two samples ($n = 429$ and $1,657$) were compared with traditional paper-and-pencil measure data from two university samples ($n = 760$ and 148) in three psychometric studies of a new measure of self-trust, the Self-Trust Questionnaire. With the exception of scale score variance, which was larger in WWW samples, psychometric properties of the scale were comparable across samples collected from the two sources. WWW samples were more similar to psychology student samples (predominantly young females) than to typical Internet users. On balance, findings indicate that the advantages of the WWW as a data source, including large heterogeneous samples, outweigh problems with data accuracy and generalizability, making the WWW an attractive source of data for researchers developing self-report personality inventories.

The World-Wide Web (WWW) increasingly is being used as a method of collecting survey and questionnaire data (Schmidt, 1997) and of conducting interactive experiments. The latter range from the collection of data with random assignment to conditions to the collection of reaction-time data (Hewson, Laurent, & Vogel, 1996). A visit to the American Psychological Society Web site (<http://psych.hanover.edu/APS/>) provides a graphic example of the growing popularity of this resource as a means of data collection.

Pros and cons associated with collecting data on the WWW have been documented (Hewson et al., 1996; Schmidt, 1997). Benefits include large sample sizes, savings in time and money, a reduction in data entry errors since subjects' data are written directly to file, and the potential for greater honesty and openness because of guaranteed anonymity and confidentiality. Problems include the possibility of multiple submissions, of incomplete responses, of unacceptable responses, and maintaining security and data integrity. Schmidt has addressed these problems and offered possible methods of resolving them through the use of sophisticated common gateway interface (CGI) programs.

Generalizability of results obtained from WWW samples to the population at large is also problematic. For instance, a recent Georgia Tech (1997) Internet survey found the typical user to be a 36-year-old white married male with a college or advanced degree, with a mean household income of \$53,000, living in the United States, and work-

ing in a computer-related, education, professional, or management field. The authors note that this profile is changing rapidly, as more people, especially women, begin to use the Internet. Thus, psychologists considering the WWW as a data source must, as a result, be concerned about the representativeness of their samples. The problem is, of course, not peculiar to the WWW. The often used undergraduate student sample also has documented generalizability problems (Sears, 1986).

In the interest of having as diversified and large an empirical base as possible for the construction and validation of a new personality inventory, the studies reported here were carried out with both undergraduate and WWW samples, which permitted systematic comparison of the two sample sources. Demographic comparisons are also made with the data reported in the Georgia Tech (1997) survey of Internet users.

The Self-Trust Questionnaire

For purposes of the first author's doctoral dissertation, a personality inventory that was designed to assess hypothesized individual differences in self-trust was developed. Self-trust was conceptualized as the capacity to accept, as a *leap of faith*, one's self-worth: "An unconditioned and unquestioning acceptance of all aspects of the self, including but not limited to self-concept, behavior, judgment, emotions and feelings, abilities, self-knowledge, intuitions, and experience" (Pasveer, 1997, p. 10). Whereas constructs such as self-esteem and self-efficacy describe the strength of self-dimensions, self-trust refers to the unconscious assumption that self is acceptable.

Self-trust is assumed to be part of the legacy of the attachment process (Ainsworth, Blehar, Waters, & Wall, 1978; Bowlby, 1969), with individual differences arising from variability in the consistency and sensitivity of caregiving provided during childhood. Items developed for the self-trust questionnaire reflect the assumption that, for those lacking in complete self-trust, self-trust issues would

This article is based in part on the first author's dissertation, conducted at the University of Calgary, 1997. We thank John Mueller for the inspiration and suggestions he provided regarding the possibilities of using the Internet and Sujan Paul, who provided technical assistance in publishing the questionnaire on the Internet. Correspondence concerning this article should be addressed to K. A. Pasveer, Department of Psychology, University of Calgary, 2500 University Drive, Calgary, AB, T2N 1N4, Canada (e-mail: kpasveer@ucalgary.ca).

register in consciousness and be apparent in responses to self-trust statements. Accordingly, items were developed that asked for agreement or disagreement with statements that were positively and negatively worded with respect to self-trust. Sample items include: "I seldom question whether I have seen something correctly," "I seldom need reassurance from others about my capabilities," "When I have to make a decision I often look to others for advice," and "I often find myself questioning my version of reality."

The development and psychometric assessment of the self-trust measure was based on three studies. Study 1 was concerned with the initial construction of the Self-Trust Questionnaire (STQ). A total of 148 conceptually derived items were assessed for face validity in the authors' research group. Through this process, the instrument was reduced to 72 items and subsequently to 20 items on the basis of univariate analyses of each item. Results throughout this report are based on the final set of 20 items. Study 2 focused on establishing the construct validity of the STQ, and Study 3 examined the psychometric properties of the scale, using a large sample.

METHOD

Subjects

Study 1 subjects included both undergraduates at a large Canadian university and respondents recruited from the WWW. WWW respondents ($N = 429$) were recruited in two ways. First, an e-mail campaign was undertaken, with 740 requests being sent primarily to relationship researchers. This mailing list was compiled from the membership directories of the International Network of Personal Relationships and the International Society for the Study of Personal Relationships. Both organizations are interdisciplinary as well as international. Professionals recruited in this way were asked to complete the questionnaire themselves and to pass the URL on to colleagues and students. Sixty-one responses and 3 weeks later, recruitment on the WWW was expanded by creating links with the American Psychological Society Web site for research on the WWW and with various popular search engines, such as Yahoo, Excite, and Lycos. Undergraduate respondents ($N = 760$) were recruited from undergraduate psychology courses.

Because Study 2 subjects were required to complete a battery of measures, including the STQ, data were gathered exclusively from students using traditional paper-and-pencil procedures. The subjects consisted of 148 graduate and undergraduate students from the same university as Study 1 and were recruited by telephone through the Department of Psychology Participant Pool and flyers posted around campus. Volunteers were given the opportunity to participate in a lottery for a \$100 main prize and two \$50 secondary prizes.

Study 3 subjects ($N = 1,657$) were recruited exclusively from the WWW, using the same links described for Study 1 but with no active solicitation. Thus, the findings from Study 3 are more representative of what might be expected from passively *posting* a questionnaire on the WWW.

Procedure

Paper-and-pencil questionnaire. Study 1 respondents received their questionnaires in class. The subjects from introductory psychology classes completed them on their own and returned them in class within 2 or 3 days. In Study 2, all the subjects completed the STQ and other measures during a scheduled laboratory session.

Electronic questionnaire. The STQ was published on the WWW in the following manner. First, respondents were required to respond to an informed consent page describing the purpose of the study and their rights as subjects. As a part of the informed consent, the subjects were told that they would not receive a score on the STQ. A button at the bottom of the page indicated that continuation with the questionnaire constituted their informed consent. Next, respondents were asked to respond to a number of demographic questions, including sex, age, marital status, country of origin, and country in which they currently resided. The questionnaire was presented next. Each item was presented with a pull-down menu that requested a response on a 5-point scale (1 = *strongly disagree*, 5 = *strongly agree*). This is the same 5-point scale as the one used for the paper-and-pencil version. Finally, on submission of the questionnaire, respondents were given more detailed information about the author's theoretical perspective on the concept of self-trust.

Data were collected and saved to file using Form Saver for Macintosh computers. No identifying characteristics, such as name or e-mail address, were requested or saved to file. The software did record the date and time of submission, the type of computer used (which often contained an individual's name), and the domain name server (DNS) of the respondent. Since this information had the potential of identifying particular individuals, the data file was cleaned weekly, and the type of computer and DNS were deleted.

This procedure was identical for Study 1 and Study 3, with the exceptions that the statement informing the subjects that no score would be given was set in bold and a larger font was used for Study 3 as a result of a problem with multiple submissions in Study 1.

RESULTS

Sample Comparisons

The demographic profile of each of the two student samples and of the two WWW samples is presented in Table 1, along with relevant demographic information from the Georgia Tech (1997) survey of Internet users.

In making sample comparisons, systematic sources of selection bias must be kept in mind for some of the samples. In Study 1, WWW subjects included relationship researchers and their acquaintances, who were contacted directly, along with self-selected Internet users. The Study 2 student sample included primarily undergraduates but some graduate students as well. Additionally, recruiting focused on people over the age of 21, and an attempt was made to optimize the proportion of males to females. Thus, for purposes of comparing a *typical* undergraduate sample of volunteers used in psychological research with a sample of entirely self-selected WWW subjects, the Studies 1 and 3 samples are most meaningful.

Sex. The ratio of females to males in the Study 1 student sample (3.54:1) is comparable to the ratio of females to male psychology majors at the same university (3.83:1) and to those awarded psychology degrees (3.22:1) in the previous academic year (1995/1996) but somewhat higher than the ratio in the Study 3 WWW sample (2.44:1). Most noteworthy is the sex ratio reversal between all of the study samples and the Georgia Tech survey ratio of .63 females to males in the general Internet user sample.

Age. Not surprisingly, when student samples are not selected on the basis of age (Study 1), they are younger than self-selected WWW samples involved in the same

Table 1
Sample Characteristics

Characteristic	Study 1		Study 2	Study 3	Georgia
	Student (<i>N</i> = 760)	WWW (<i>N</i> = 429)	Student (<i>N</i> = 148)	WWW (<i>N</i> = 1,657)	Tech Survey (<i>N</i> = 10,108)
Sex					
Male	22%	38%	35%	29%	61%
Female	78%	62%	65%	71%	39%
Age					
Overall range	17–54	13–62	18–54	11–78	5–80
Under 16 or age unknown	1%	4%	2%	7%	2%
16–20	59%	24%	14%	29%	9%
21–25	26%	27%	52%	25%	14%
Over 25	14%	45%	32%	39%	75%
<i>M</i>	21.9	27.5	25.5	25.9	35.7
<i>SD</i>	6.5	9.7	6.9	9.4	
Marital Status					
Single	85%	64%	69%	68%	38%
Married	8%	28%	16%	22%	40%
Common law	7%	8%	15%	10%	9%
Education					
College or advanced degree			12%	35%	47%
Some college	100%		88%	41%	28%
High school or tech/trade				18%	18%
Less than H.S.				3%	2%
Residence					
Canada/United States	100%	78%	100%	83%	90%
Europe		11%		9%	7%
Other		11%		8%	3%

research [Study 3; $t(2,273) = -12.18, p < .0009$] and also less variable in age [$F(1,2273) = 145.31, p < .0009$]. However, all of the research samples were, on average, younger than typical Internet users.

Marital status. The subjects in the self-trust research, particularly in the student sample not selected on the basis of age, were predominantly single. This is in contrast to the general Internet user population, where a somewhat larger proportion were in a continuing close relationship than were single. This is not surprising, given the younger age of the student sample.

Education. Inasmuch as both student samples are limited to university students, education level comparisons with WWW samples are not meaningful. Comparison of the Study 3 WWW sample and the survey of Internet users indicates that, although in both cases the overall level of education is high (the vast majority have at least some college or university), the proportion with completed college degrees was somewhat higher in the Georgia Tech sample.

Residence. Residence comparisons, on the face of it, would suggest that the WWW samples were somewhat more heterogeneous than the student samples. However, residence by itself is a poor proxy for heterogeneity on potentially important dimensions, such as family income level, ethnic background, and religious background. Students may reside in the country of the university they attend but come from a wide range of backgrounds. Similarly, the heterogeneity within the Canadian and the American WWW users may be as large or larger than differences between Canada and the United States and other parts of the world.

Summary. The subjects in the self-trust research, whether students completing paper-and-pencil questionnaires or WWW subjects completing the inventory electronically, tended to be predominantly female, young, single, well educated, and, overall, more similar than different. Contrasts were most striking between the research samples and the typically male and somewhat older general Internet user who was, more often than not, in a continuing close relationship.

Self-Trust Comparisons

Psychometric data for the self-trust scale in the four research samples are summarized in Table 2.

Descriptive analyses. Paired comparison (Tukey HSD) analyses of the mean scale scores revealed only one reliable difference: paper-and-pencil subjects in Study 2 had a higher mean item scale score than Study 1 paper-and-pencil subjects ($p < .03$). This difference may be of

Table 2
Descriptive and Psychometric Analyses of the STQ Samples

Measure	Study 1		Study 2	Study 3
	Student (<i>N</i> = 760)	WWW (<i>N</i> = 429)	Student (<i>N</i> = 148)	WWW (<i>N</i> = 1,657)
Item mean	3.24*	3.26*†	3.40†	3.29*†
Standard deviation	0.56*	0.63†	0.60*†	0.73‡
Mean interitem correlation	.21	.23	.23	.27
Cronbach's alpha	.84	.86	.86	.88

Note—Values sharing a common superscript are not significantly different ($p > .05$). Mean comparisons are based on Tukey (HSD) test, and standard deviation comparisons are based on Levene's test for equality of variances.

theoretical significance to the extent that the paper-and-pencil sample in Study 1 was younger, with a higher proportion of females. However, there were no consistent sex or age differences across the studies. More importantly, at about one quarter of a standard deviation, it is likely that the difference (0.16) is of little practical significance.

Comparison of sample variances indicated that the WWW samples were more variable than the paper-and-pencil samples, particularly when the WWW sample was large. To assess the extent to which differences in variability reflect underlying differences in the heterogeneity of the samples, a sample from the Study 3 WWW sample was matched on gender, age, education, marital status, and residence with the 148 subjects in Study 2. As 3 subjects in Study 2 failed to report age or gender, the final analysis was based on 145 matched respondents.

In the matched comparison, the mean item score for students ($M = 3.39$) did not differ from the WWW sample ($M = 3.26$). However, even in the matched samples, there was more variability in the WWW group ($SD = 0.73$) than in the student group ($SD = 0.59$), as assessed by Levene's test for equality of variances [$F(1,144) = 8.17, p < .005$]. This difference in variability may reflect other sources of uncontrolled heterogeneity differences between the two samples or may indicate response bias differences between the paper-and-pencil and electronic formats. For instance, using a pull-down menu for the response scale on the WWW may have made it more difficult for the subjects to track their responses across items and, hence, sustain patterns of consistency in their responses, than was the case for paper-and-pencil subjects.

Psychometric analyses. As shown in Table 2, measures of internal consistency for the STQ were very similar for WWW and student samples. The factor structure of the STQ was also very similar in factor analyses of the scale in each sample. In each case, four factors assessing different theoretically interpretable aspects of self-trust emerged: general self-trust, external validation, experiential trust, and emotional trust.

Summary

But for an apparent tendency for more variance in WWW scores, data gathered from the WWW yielded descriptive and psychometric results that were remarkably similar to data gathered using paper-and-pencil procedures from student samples.

GENERAL DISCUSSION

Use of the WWW as a tool for data collection facilitated the collection of data in all phases of questionnaire construction. Large samples were collected from diverse populations, with, for instance, educational levels ranging from less than a high school degree to a Ph.D. This method of data collection allowed us to have some confidence that the psychometric properties of a new personality inventory generalize beyond the typical university

population of late adolescents to a more geographically dispersed population of computer-literate adults. In this instance, we found the student sample to be comparable with the more heterogeneous sample with respect to the basic properties of the scale. Thus, psychometric researchers looking for large samples with which to study the correlates and structure of measures would do well to consider using the WWW as a data source. In the present project, the large sample benefits are clearer than the generalizability issues, insofar as both sample sources were disproportionately female, well educated, young, and likely to be single. Thus, researchers may find more heterogeneity in their WWW samples but still be left with important unresolved generalizability issues.

Sample demographic comparisons also indicate that specific WWW samples need not fit the profile of the prototypical Internet user as a young professional male (Georgia Tech, 1997). In all three STQ studies, females significantly outnumbered males, possibly because the research topic is of greater interest to females. In addition, the mean age of WWW samples obtained in this investigation was younger than that obtained by Georgia Tech, and the subjects were predominantly single rather than married. The general level of education appeared to be comparable.

It is noteworthy that the common concern about the unrepresentativeness of university student samples (Sears, 1986) was not borne out in this investigation. Although it must be stressed that the Internet samples are not representative of the general population, there was evidence that the student samples were at least comparable with this computer-literate population.

Other advantages were the ability to collect large amounts of data with a minimum of time and effort expended and savings in time and data entry errors by having the data written directly to file.

Drawbacks

The major difficulty encountered arises from the respondents' propensity to submit the questionnaire more than once. On one occasion, a subject submitted the questionnaire a total of 10 times. These multiple submissions were found and duplicate data deleted, when the forms were submitted in close time proximity. This type of multiple submission can be caught and immediately corrected with a sophisticated CGI program (Schmidt, 1997).

For researchers who do not have access to such programs, it is necessary to look for the possible causes of multiple submissions. One possibility, in this case, may be that these respondents expected feedback in the form of a score. When that was not forthcoming, they may have resubmitted the form, thinking that an error had occurred in the program. Although the informed consent clearly stated that subjects would not receive scoring information, many may not have read the informed consent in its entirety. If this is the case, the problem can be easily corrected by having a screen appear on submis-

sion of the questionnaire that both thanks the respondent and explains why a score will not be forthcoming.

It is possible that other respondents submitted the form more than once and that this was not caught. This would have been the case if the forms were submitted hours or days apart, during which time other information was submitted. Because of ethical guidelines prohibiting any type of identifying characteristics, the DNS (identification of the server) was the only data submitted that could identify the respondent. This was deleted regularly, and thus it was not possible to screen data for multiple submissions that did not immediately follow one another. Again, a more sophisticated data collection program (Schmidt, 1997) possibly could have caught and resolved problems such as this by locating forms that had identical demographic information and were very similar in response. Unfortunately, this solution would also potentially eliminate valid response sets (Schmidt, 1997). Even retention of the DNS entry in the data file will not aid in locating these types of duplicates, since thousands of people may respond from the same domain. Thus, unless duplicate entries are made in close proximity in time, there is not yet a satisfactory solution to this problem.

Subjects may also have given false information on some of the demographic questions. For example, at least one 16-year-old subject indicated that he or she had obtained a Ph.D. Some time and effort are needed to locate and delete obviously erroneous data such as these. Other false information, such as incorrect age, gender, and so forth, is not amenable to detection. This particular problem is not unique to Internet data collection, however. Anytime subjects are given forms to fill out, this potential exists. The Internet, in fact, may be *less* likely to elicit erroneous responses. Since responses are obviously anonymous and confidential, respondents may be more honest with their answers (Hewson et al., 1996). The best we can do as researchers is to inspect responses and eliminate any answers that are obviously incorrect.

On balance, data collection through use of the WWW was successful. Active recruitment of subjects was not necessary, and data were collected in a form suitable for data

analysis. Although drawbacks were encountered, we believe the advantages of a large sample size justify using the WWW as a data collection source. Having entirely *clean* data sets is always a challenge, but the virtual impossibility of checking the veridicality of data on the WWW requires researchers to use data from the Internet cautiously. Procedures for minimizing inaccurate or fraudulent data reporting should involve, at minimum, including sufficient demographic questions to catch transparent inconsistencies (e.g., 16-year-old PhD holders) and studies run in parallel with non-WWW samples in order to corroborate conclusions drawn from WWW-based data. In order to more fully exploit the WWW as a data collection resource, researchers will need to explore additional mechanisms for checking data veridicality. The effort involved in addressing the data accuracy issue will be well spent, given the potential size and diversity of WWW samples.

REFERENCES

- AINSWORTH, M. D. S., BLEHAR, M. C., WATERS, E., & WALL, S. (1978). *Patterns of attachment: A psychological study of the strange situation*. Hillsdale, NJ: Erlbaum.
- BOWLBY, J. (1969). *Attachment and loss: Attachment* (2nd ed.). New York: Penguin.
- GEORGIA INSTITUTE OF TECHNOLOGY: GEORGIA TECH'S GRAPHICS, VISUALIZATION, AND USABILITY CENTER (1997). *GVU's WWW User Surveys* [On-line]. Available URL: http://www.cc.gatech.edu/gvu/user_surveys/
- HEWSON, C. M., LAURENT, D., & VOGEL, C. M. (1996). Proper methodologies for psychological and sociological studies conducted via the Internet. *Behavior Research Methods, Instruments, & Computers*, **28**, 186-191.
- PASVEER, K. A. (1997). *Self-trust: Definition and creation of the self-trust questionnaire*. Unpublished doctoral dissertation, University of Calgary.
- SCHMIDT, W. C. (1997). World-Wide Web survey research: Benefits, potential problems, and solutions. *Behavior Research Methods, Instruments, & Computers*, **29**, 274-279.
- SEARS, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality & Social Psychology*, **51**, 515-530.

(Manuscript received January 6, 1998;
revision accepted for publication March 4, 1998.)