Fordham University
DigitalResearch@Fordham

Psychology Faculty Publications | Psychology

1-1-2003

Validating Internet Research: A Test of the Psychometric Equivalence of Internet and In-Person Samples

Paul Meyerson
Fordham University

Warren W. Tryon
Fordham University

Follow this and additional works at: http://fordham.bepress.com/psych_facultypubs

Part of the Psychology Commons

Recommended Citation
Meyerson, Paul and Tryon, Warren W., "Validating Internet Research: A Test of the Psychometric Equivalence of Internet and In-Person Samples" (2003). *Psychology Faculty Publications.* Paper 22.
http://fordham.bepress.com/psych_facultypubs/22

This Article is brought to you for free and open access by the Psychology at DigitalResearch@Fordham. It has been accepted for inclusion in Psychology Faculty Publications by an authorized administrator of DigitalResearch@Fordham. For more information, please contact considine@fordham.edu.

Behavior Research Methods, Instruments, & Computers
2003, 35 (4), 614-620

Validating Internet research: A test of the psychometric equivalence of Internet and in-person samples

PAUL MEYERSON and WARREN W. TRYON
Fordham University, Bronx, New York

This study evaluated the psychometric equivalency of Web-based research. The Sexual Boredom Scale was presented via the World-Wide Web along with five additional scales used to validate it. A subset of 533 participants that matched a previously published sample (Watt & Ewing, 1996) on age, gender, and race was identified. An 8 × 8 correlation matrix from the matched Internet sample was compared via structural equation modeling with a similar 8 × 8 correlation matrix from the previously published study. The Internet and previously published samples were psychometrically equivalent. Coefficient alpha values calculated on the matched Internet sample yielded reliability coefficients almost identical to those for the previously published sample. Factors such as computer administration and uncontrollable administration settings did not appear to affect the results. Demographic data indicated an overrepresentation of males by about 6% and Caucasians by about 13% relative to the U.S. Census (2000). A total of 2,230 participants were obtained in about 8 months without remuneration. These results suggest that data collection on the Web is (1) reliable, (2) valid, (3) reasonably representative, (4) cost effective, and (5) efficient.

Both experimental and survey research can now be conducted on the World-Wide Web (WWW). The most recent and thorough discussion of behavioral research on the Internet has come in the form of two books by Birnbaum (2000a, 2000b) and one by Batinic, Reips, and Bosnjak (2002). Reips (2001) has reviewed the history of Web experiments and described 5 years of experience with an Internet research resource for conducting true experiments on the WWW at http://www.psychologie.unizh.ch/genpsy/Ulf/Lab/WebExpPsyLab.html and http://www.genpsylab.unizh.ch/wextor/index.html. Gordon and Rosenblum (2001) have described an audiovisual speech Web lab suitable for teaching and research at http://www.psych.ucr.edu/avespeech/lab. Schmidt (1997) has announced the availability of a utility program called the WWW Survey Assistant to facilitate the creation of an HTML (hypertext markup language)–CGI (common gateway interface) document for administering surveys on the WWW at http://or.psychology.dal.ca/~wcs. No HTML or CGI programming skills are needed. Carlton et al. (1999) have described a Web-based digitized video image system for the study of motor coordination (cf. http://www.iines.uiuc.edu/digi-net97/survey/).

A fundamental assumption of Internet (on-line) research is that the results obtained are comparable to in-person (off-line) research. Investigators have taken different approaches to this matter, but none have specifically examined psychometric equivalence. Here findings of other investigators will first be briefly reviewed and psychometric equivalence then defined.

Krantz, Ballard, and Scher (1997), Smith and Leigh (1997), Pasveer and Ellard (1998), Stanton (1998), Buchanan and Smith (1999a, 1999b), Davis (1999), and Buchanan (2000) have reported similar means and standard deviations for Web-based and in-person studies. Pasveer and Ellard reported similar results for the Self-Trust Questionnaire on Internet and in-person samples.

Buchanan and Smith (1999b) reported comparable coefficient alphas and confirmatory factor structures for Internet and in-person administration of the Self-Monitoring Scale–Revised (SMS–R: Buchanan & Smith, 1999b; Snyder, 1974; Snyder & Gangestad, 1986). Stanton (1998) reported a similar factor structure for an organizational justice scale when Internet and in-person data were compared. Davis (1999) reported comparable coefficient alphas across Internet and in-person samples responding to the Ruminative Responses Scale (RRS) of the Response Styles Questionnaire (RSQ; Nolen-Hoeksema & Morrow, 1991). Their results were consistent with in-person data published by Snyder, Berscheid, and Glick (1985) and Buchanan (2000).

Cronk and West (2002) examined the comparability of Web and in-person data using a 2 (Web vs. paper-and-pencil) × 2 (in class vs. outside class) analysis of variance (ANOVA) design with statistical power of .94 and did not find any significant differences. Buchanan and Smith (1999a) predicted and observed that Usenet Newsgroup users are likely to be high or low self-monitors.

Krantz and Dalal (2000) have summarized research regarding the validity of WWW-based studies. They con-

Correspondence concerning this article should be sent to W. W. Tryon, Department of Psychology, Fordham University, 441 East Fordham Road, Bronx, NY 10458-5198 (e-mail: wtryon@fordham.edu).

Copyright 2003 Psychonomic Society, Inc.

clude that "In all cases, there seems to be a surprising match between laboratory and WWW versions of surveys, scales, and experimental variables" (p. 56). They further conclude in connection with their own research that off-line and on-line research from the same study yield results that can "essentially replace each other" (p. 56).

No one has previously demonstrated the psychometric equivalence of Internet and in-person data. As Honaker (1988) has stated, "psychometrically, two forms of a test are considered equivalent if it has been demonstrated that the two forms are *parallel*" (p. 562). Ghiselli (1964, p. 227; Ghiselli, Campbell, & Zedeck, 1981, pp. 192–227) has identified the following three requirements as necessary for establishing parallel forms: (1) equal means, (2) equal variances, and (3) correlate to the same degree with other variables. If two test forms meet Criteria 2 and 3 but have different means, they can be made parallel by adding or subtracting a constant. Documenting mean differences is an important part of empirical research and hypothesis testing. The point of interest here is that if Internet and in-person results were shown to differ always by a fixed amount, then adding or subtracting a constant would be sufficient to make the two data sets psychometrically equivalent. Hence, demonstrating mean differences is not in and of itself sufficient grounds for concluding that two tests cannot be made parallel. For this reason, mean differences are not tested for in the results presented below.

If two test forms meet Criteria 1 and 3 but have different variances, they can be made parallel through an equipercentile transformation. Documenting differences in variation is also an important part of empirical research and hypothesis testing as illustrated by the many studies based on ANOVA. The main point here is that if Internet and in-person results were shown to differ only in terms of variation, they could be made to have the same variance through an equipercentile transform and therefore made to be psychometrically equivalent.

The third requirement that the two tests, or forms of a test, be highly correlated with each other is crucial; it is a necessary condition for psychometric equivalence. Construct validity concerns the correlations of the target test with other theoretically pertinent variables. Psychometric equivalence is therefore most clearly demonstrated when the correlation matrix of the first test with validating variables is highly similar to the correlation matrix of the second test with the same validating variables. A multivariate comparison of the two correlation matrices is the most efficient and certain way to test for psychometric equivalence. Factor analyses and structural equation models are not more informative with regard to psychometric equivalence, because they derive from the correlation (covariance) matrices that have already been compared.

Several of the studies reviewed above indirectly addressed psychometric equivalence by comparing factor structures and even more indirectly addressed psychometric equivalence by comparing coefficient alphas. No one has yet directly determined whether the correlation matrices derived from Web-based research are comparable to those obtained during in-person research according to the criteria for psychometric equivalence specified above. Our purpose in the present research was to evaluate the psychometric equivalency of Internet versus in-person research by comparing an entire correlation matrix of a target test and its validation scales administered with previously collected in-person data while controlling for demographic characteristics. Replicating results from a previously published report of in-person data using the Internet was deemed to be the strongest test of psychometric equivalence. We chose to use Watt and Ewing's (1996) data, for reasons provided in the Instruments section below.

METHOD

Participants

A total of 2,230 Internet participants were obtained from which a subsample of 533 participants matched on age, gender, race, and number of sexual relationships was culled. This sample is adequate to support the LISREL (Jöreskog & Sörbom, 1996) analyses reported below. Matching to the 253 participants in the previously published validation study (Watt & Ewing, 1996) was done on a group rather than an individual basis. Table 1 presents the gender, ethnicity, and age breakdown for the previously published and matched Internet groups, all Internet participants (complete Internet sample), and the U.S. Census (2000) comparison group. Matching was evaluated by t test except in comparisons made with the census group, because standard deviations were not available for the U.S. Census data. The matched group was not statistically different from the group published by Watt and Ewing (1996) on any of the demographic characteristics by design. The proportion of males in the matched Internet group (.4109) was significantly lower than the proportion of males in the U.S. Census (.4880) ($Z = -3.60$, $p < .001$). The proportion of Caucasians in the matched Internet group (.8480) was significantly greater than the proportion of Caucasians in the U.S. Census (.7137) ($Z = -6.85$, $p < .001$). The proportion of males in the complete Internet group of 2,230 participants (.5502) was significantly greater than the proportion of males in the U.S. Census (.4880) ($Z = 5.79$, $p < .001$) and the proportion of Caucasians in the complete Internet group (.8422) was also significantly greater than that in the U.S. Census (.7137) ($Z = 13.42$, $p < .001$). It appears that white males remain overrepresented in Internet research samples. All participants in the previously published validation sample had at least one sexual relationship, and 71% were currently in a relationship. All participants in the matched Internet sample had at least one sexual relationship, and 63% were currently in a relationship. As in the original study, any participants not reporting any sexual history were excluded from the results.

Participants had to click through an age consent form before gaining access to the questionnaire s, because of their sexual themes. The form presented the following two hyperlink options: (1) Are you at least 18 years old? And (2) Are you under 18 years old? Only those that clicked on the proper link could enter the Web site. All others were taken to http://www.yahoo.com. Although clearly age can be faked, this type of simple disclaimer is frequently used by adult-oriented Web sites, and it was approved by the Fordham IRB. APA's "Psychological Research on the Net" page,

Table 1
Demographic Data for the Previously Published (Watt & Ewing, 1996) Participants, Matched Internet Participants, All Internet Participants, and U.S. Census Data for the Year 2000

	Watt & Ewing (N = 253)		Matched Internet (N = 533)		Complete Internet (N = 2,230)		Census Data (N = 275,843,000)	
	n	%	*n*	%	*n*	%	*n*	%
				Gender				
Male	84	33.20	219	41.09	1,227	55.02	134,871,000	48.89
Female	169	66.80	314	58.91	1,003	44.98	140,972,000	51.11
				Ethnicity				
Caucasian	235	93	452	84.80	1,878	84.22	196,875,000	71.37
Noncaucasian	19	7	81	15.20	352	15.78	78,968,000	28.63
				Age				
Age (*M*)	20.9		21.35		28.83		36.5	
Age (*SD*)	4.42		2.58		10.37		n/a	

http://psych.hanover.edu/ Research/exponnet.html, contains links that employ this type of click-through consent form to select adult-only populations (see http://www.geocities.com/ Athens/Oracle/ 3900/psychinfo.html, http://www.liii.com/~fantine/consent.html, or http://www.cwru.edu/artsci/pscl/forgive/main.html for examples), and no superior practical alternative was available.

In addition to the entrance consent form, all participants had to "click" on a button that gave their consent before entering the study. Although through use of bookmarks in a Web browser it could have been possible to enter the study without first accessing the consent form, the site was set up in such a way that data were not accepted as valid without the subject's having started at the consent form and preceded sequentially through the tests.

Instruments

Test selection entailed applying the following criteria to an extensive PsycInfo on-line search of approximately 4,500 tests. First, the target measure must have been conducted on a nonpsychiatric population. Although at some point it may become possible to recruit on line from psychiatric or other specific populations of interest, it does not make sense to try this before knowing whether the Internet can provide valid data with normal people. Second, the target study must not have a test–retest or an interview requirement. Although these requirements should be relatively easy to meet in the future by means of unique identifiers or live "chat" sessions, they presented unnecessary complexities for the present study. Third, the tests used to validate the target test must also fit the second criterion above. Finally, permission to reproduce the tests on line must be obtained from either the author or the publisher, and in most cases both. A total of 30 tests met these criteria. Some studies used either too few measures or required some type of direct interaction with the subjects. Ultimately, eight measures fit the first three criteria and had a manageable number of validation tests. The final limiting factor was permission to reproduce the target measure and all of the supporting measures. Permission was granted only for the Sexual Boredom Scale and its validation scales (Watt & Ewing, 1996) described next.

The Boredom Proneness Scale (Farmer & Sundberg, 1986) is a 28 item true–false, self-report scale. It was modified from the previously published Sexual Boredom Scale validation study from the true–false format to a Likert scale ranging from 1 (*highly disagree*) to 7 (*highly agree*), with higher total scores indicating higher boredom proneness.

The Index of Sexual Satisfaction (Hudson, Harrison, & Crosscup, 1981) is a 25-item scale on which the participant rates each question from 1 (*rarely or none of the time*) to 5 (*most or all of the time*). High scores are said to represent "the presence of a sexual problem." The Satisfaction With Life Scale (Diener, Emmons, Larsen, & Griffin, 1985) is a 5-item questionnaire scored on a Likert scale, where 1 indicates *low satisfaction* and 5 indicates *high satisfaction*.

The concept of sensation seeking as an aspect of sexual boredom was assessed via form V of the Sensation Seeking Scale (Zuckerman, 1979). This scale consists of 40 forced-choice items relating to four factors: Thrill and Adventure Seeking (desire to engage in physical risk-taking activities), Experience Seeking (desire to pursue new experiences through the mind and senses), Disinhibition (desire to disinhibit oneself in social situations in the pursuit of pleasure), and Boredom Susceptibility (reflects one's tendency to get bored and restless without novel stimulation).

The 30-item Sexuality Scale (Snell & Papini, 1989) was used to assess "Sexual Esteem (I am better at sex than most other people)," "Sexual Depression (I am disappointed about the quality of my sex life)," and "Sexual Preoccupation (I think about sex all the time)."

Procedures

All tests were administered via the World-Wide Web. The Web site was set up by creating the required files and then sending them, via file transfer protocol (FTP), to a UNIX server at an Internet service provider (ISP). These HTML files are accessible to anyone in the world with a suitable browser. In order to publicize the site, it was submitted to a search engine submission tool, http://uswebsites.com/submit/, which dispersed the uniform resource locator (URL) of the Web site to 30 different search engines. Two other search engine submission tools were also used: http://www.netcreations.com/postmaster/, and http://www.register-it.com/. The page was also submitted to a list of search engines available at http://www.virtualpromote.com/ promoteb.html.

The Web site was available at http://www.sexualboredom.com. This page consisted of a welcome page that presented the study as helping to develop a measure of sexual boredom and stating that interested parties should continue to the next page to read the consent form. There was a link to the author's and the dissertation mentor's e-mail addresses on every page of the study. The home page also requested other Web site administrators to link the present site to theirs in an attempt to increase the number of hits the site would get. Finally, the site had an "18 or older" set of links, one leading to the consent page. The Fordham University IRB approved this procedure.

The informed consent page was the same that is currently used for in-person research. The consent page also had two buttons, one labeled "I wish to participate" and one labeled "I decline to participate." Agreeing to participate took the participant to the demographic page. The Demographics page asked for age, ethnicity, religion, level of education, SES, and gender formatted as "radio buttons"; that is, all of the choices for this information were pres-

ent, and the participants had to pick the one that best fit them. Age and SES required participants to type a number. This was done to avoid arbitrary stratification where no clear groups exist. Participants then clicked on either a *reset information* or a *submit information* button. The submit button brought them to the body of the survey.

The survey pages consisted of one long page for each of six tests. Each survey was presented in a fashion similar to that for the demographic data, in that there were "radio buttons" that had the relevant scoring options (i.e., one labeled button for each response option). Submission of one scale automatically brought up the next scale. If the subject did not answer all of the questions, he or she was prompted to do so after pressing *submit*. No tests with missing data could be submitted. This "error checking" is another benefit of computerized testing, including Web testing.

After all of the tests had been completed, the subject was sent to a summary page where the results of their scales were compared with those of the previously published, validation population for informational purposes only. This positive feedback rewarded the participants for participating in the research without giving any serious clinical feedback. This page reported the participant's total Sexual Boredom Scale score and noted whether it was higher, lower, or the same as the Watt and Ewing (1996) average score. Participants were given the previously published demographic information and told that the results were provided strictly for entertainment. There was a *continue* button from this page, as well as the same e-mail links to ask the researchers questions as were present on all prior pages. This page was made accessible only if one submitted all of the required scales. The continue button linked the participant to the debriefing page. No other page linked to the debriefing page, and it was not possible to go directly to that page without knowing the exact page address, which was not publicized. The participants were told that this study was a replication of a prior study for the purposes of determining the differences, if any, that should be expected between a Web-based survey and a "real-life" survey. They were reminded that the information they provided was confidential and anonymous. There was an e-mail link to both authors for questions or comments, and there was a link to a *results* page.

The results page permitted the participants to input their e-mail addresses to have an abstract of the study e-mailed to them upon completion. This information was *not* linked to the participant's survey responses, and this fact was made clear. The results page stood alone, and the information on it was not connected to the response data whatsoever.

RESULTS

Four participant groups will be discussed. Henceforth, the term *Watt and Ewing sample* refers to the group of 253 participants that took part in the previously published validation study done by Watt and Ewing (1996). The term *matched sample* refers to the subset of 533 Internet participants that matched the demographic characteristics of the previously published group. The term *complete Internet sample* refers to all 2,230 Internet participants. The term *census data* refers to the projected 2000 population (275,843,000) of the United States of America.

Sample Size

The 2,230 Internet participants who completed all 145 questions across all tests were obtained over approximately 9 months, representing an average of approximately 9 data sets per day. The Web site received approximately 12,000 hits over that time. The ratio of number of completed studies to number of total hits is .1979, or approximately 20%. More than one third of all participants who completed this study were referred by one of the major search engines, which did not begin referring participants until almost 6 months after the study began.

Descriptive Statistics

Table 2 contains the mean, minimum, and maximum scores, with standard deviations, for all scales and subscales administered for the previously published group, the matched Internet group, and the complete group of all Internet participants. Means were not compared across groups; such differences do not bear on psychometric equivalence since any of them can be completely nullified by adding a constant.

Table 3 presents the coefficient alpha values for each of the independent scales in the previously published and the present study. The matched group alphas are the alpha values calculated on the present Internet-based results. Numerical equivalence when rounded to the first decimal is more frequently found than not. Deviations are confined to ±.1, indicating that the Internet data are as reliable as paper-and-pencil data.

Inferential Statistics

Test of matrix equivalence. Psychometric equivalence requires similar correlations, as previously described, because means and variances can be equated through transformations. LISREL (Jöreskog & Sörbom, 1996) requires that the correlation matrices being compared are positive and definite. The correlation matrix of the 13 variables presented in Table 2 was not positive definite. This can occur because of colinearity, where one variable is fully predictable from another or combination of other variables. Deletion of the Sensation Seeking Scale total score and the Sexuality Scale total score plus its subscales finally produced positive definite correlation matrices. Gender was not included, because it was a noncontinuous, binary variable and therefore did not form true Pearson product–moment correlations with the other variables. The final 8 × 8 correlation matrix for the previously published sample is presented in Table 4. The corresponding 8 × 8 correlation matrix for the matched Internet sample is presented in Table 5. These two matrices were compared using LISREL 8 (Jöreskog & Sörbom, 1996).

Goodness-of-fit indices. The chi-squared noncentrality parameter index (CSNCP) is a measure of the overall fit of the model to the data. In this case, it quantifies the difference between the two correlation matrices. Smaller values indicate a better fit (Jöreskog & Sörbom, 1996). Chi-square values less than 100 usually suggest a good fit. In the present study, CSNCP = 14.09. The 90% confidence interval ranged from 0.66 to 35.46. Even the top of the confidence interval is well below 100. The root mean square error of approximation (RMSEA) is based

Table 2
Descriptive Statistics for the Previously Published (Watt & Ewing, 1996) Sample (*N* = 253), Matched Internet Sample (*N* = 533), and all Internet Participants (*N* = 2,230)

	Watt & Ewing Sample				Matched Internet Sample				Complete Internet Sample			
Scale	*M*	*SD*	Min	Max	*M*	*SD*	Min	Max	*M*	*SD*	Min	Max
Sexual Boredom	48.00	18.62	20.00	105.00	60.51	21.71	19.00	120.00	61.97	21.43	11.00	126.00
Boredom Proneness	96.12	19.39	50.00	147.00	119.77	19.18	28.00	196.00	118.31	18.67	28.00	196.00
Index of Sexual Satisfaction	47.16	13.08	25.00	97.00	54.23	16.17	25.00	113.00	57.43	17.15	25.00	125.00
Satisfaction with Life	23.69	6.79	5.00	35.00	22.36	7.16	5.00	35.00	21.54	7.31	5.00	35.00
Sensation Seeking	21.49	6.59	1.00	35.00	23.49	6.13	2.00	39.00	22.00	6.38	2.00	39.00
Thrill & Adventure Seeking*	7.42	2.46	0.00	10.00	7.04	2.57	0.00	10.00	6.37	2.75	0.00	10.00
Experience Seeking*	5.34	2.34	0.00	10.00	6.56	2.12	0.00	10.00	6.40	2.09	0.00	10.00
Disinhibition*	5.48	2.77	0.00	10.00	5.73	2.21	1.00	10.00	5.33	2.26	0.00	10.00
Boredom Susceptibility*	3.21	2.04	0.00	9.00	4.16	2.24	0.00	10.00	3.90	2.16	0.00	10.00
Sexuality Scale†,‡	−4.95	13.14	−45.00	38.00	−2.51	12.42	−41.00	40.00	−5.55	13.03	−45.00	42.00
Sexual Depression‡	−9.77	7.74	−20.00	10.00	7.77	8.87	−20.00	20.00	2.85	7.40	−20.00	20.00
Sexual Esteem‡	−8.27	7.84	−16.00	20.00	−8.03	8.59	−20.00	20.00	−4.36	6.34	−20.00	19.00
Sexual Preoccupation‡	−1.03	8.30	−17.00	16.00	−4.70	9.71	−20.00	20.00	−4.98	8.66	−20.00	20.00

*Subscales of Zuckerman's (1979) Sensation Seeking Scale. †Subscales of Snell and Papini's (1989) Sexuality Scale. ‡Items are rated on a −2 (*disagree*) to +2 (*agree*) scale.

on the chi-squared noncentrality parameter (NCP) and measures the degree to which the model fails to fit the data accurately. It is a "measure of discrepancy per degree of freedom" (Browne, 1993). Values close to 0.05 indicate a good fit. The present study had an RMSEA value of 0.036, with the 90% confidence interval ranging from 0.008 to 0.058. The goodness-of-fit index (GFI) is a measure that compares how much better the data fit an ideal model than no model at all. The expected range for positive results here is between 0.90 and 1.00, with 1.00 representing a perfect data fit. The results in the present study were GFI = 0.97. The normed fit index (NFI) and the nonnormed fit index (NNFI) quantify how much better one model fits in comparison with a baseline model (Jöreskog & Sörbom, 1996). They should be between 0.95 and 1.0. In the present study, NFI = 0.96 and NNFI = 0.97. The expected cross-validation index (ECVI) controls for the number of parameters because it is always possible to generate a close fit by including nonessential parameters. The ECVI gives results for a perfectly saturated model and for a worst-possible-fit independence model. In the present study, the perfectly saturated model gave an ECVI score of 0.092. The completely independent model yielded an ECVI of 1.38. The present results were ECVI = 0.17 with a 90% confidence interval of 0.15 to 0.19, which is rather close to the perfectly fitting saturated model.

DISCUSSION

The data presented above indicate that the reliability (internal consistency) of Internet data is comparable to the reliability of in-person data and that Internet validity coefficients are psychometrically equivalent to those obtained from in-person data. These results support the reliability and validity of Internet testing and add to a growing body of evidence reviewed above that the Internet is a valid forum for conducting psychological research.

Modest participant variability has been investigated. Bailey, Foote, and Throckmorton (2000) reported greater ethnic diversity than did Watt and Ewing (1996), although the former sample came from California, whereas the latter came from Kansas. The average age for the previously published group was fairly young, but not unusual for a college sample. Both samples were considerably younger than the U.S. Census (2000) average. Threats to the external validity of research results based on undergraduate samples of convenience have long been acknowledged but the practice of conducting research with these unrepresentative samples has continued largely for lack of practical alternatives. The Internet promises a substantially more diverse participant pool. However, complete parity with the U.S. Census has yet to be achieved. Caucasians and men were somewhat overrepresented in our complete Internet sample of 2,230 participants, but the difference in proportions was not par-

Table 3
Coefficient Alpha for the Sexual Boredom Scale and its Validation Scales for the Watt and Ewing (1996) Sample (*N* = 253) and the Matched Internet Sample (*N* = 533)

Measure	Watt & Ewing	Matched Internet
Sexual Boredom Scale	.92	.92
Boredom Proneness Scale	.79	.80
Index of Sexual Satisfaction	.91	.90
Satisfaction with Life Scale	.87	.88
Sensation Seeking Scale		
Thrill and Adventure Seeking	.76	.77
Experience Seeking	.66	.61
Disinhibition	.76	.61
Boredom Susceptibility	.56	.62
Sexuality Scale		
Sexual Depression	.90	.87
Sexual Esteem	.92	.90
Sexual Preoccupation	.88	.92

Table 4
Correlation Matrix from Watt and Ewing (1996) Sample of 253 Participants

	Sexual Boredom	Boredom Proneness	Index of Sexual Satisfaction	Satisfaction with Life	Boredom Susceptibility	Thrill & Adventure Seeking	Experience Seeking
Boredom Proneness	.39‡						
Index of Sexual Satisfaction	.33‡	.34‡					
Satisfaction With Life	−.21†	−.49‡	−.44‡				
Boredom Susceptibility	.37‡	.28†	.09	−.12			
Thrill & Adventure Seeking	.12	−.08	−.10	.15*	.16*		
Experience Seeking	.13*	−.04	−.11	−.13*	.16*	.32‡	
Disinhibition	.42†	.22†	.00	−.08	.49†	.31‡	.32‡

*$p < .05$. †$p < .01$. ‡$p < .0001$.

Table 5
Correlation Matrix from Present Internet Sample of 533 Matched Participants

	Sexual Boredom	Boredom Proneness	Index of Sexual Satisfaction	Satisfaction with Life	Boredom Susceptibility	Thrill & Adventure Seeking	Experience Seeking
Boredom Proneness	.36‡						
Index of Sexual Satisfaction	.43‡	.20‡					
Satisfaction With Life	−.22‡	−.18‡	−.33‡				
Boredom Susceptibility	.42‡	.28‡	.12†	−.17†			
Thrill & Adventure Seeking	.05	.06	−.12†	.07	.14†		
Experience Seeking	.13†	.09*	−.07	−.07	.30‡	.27‡	
Disinhibition	.31‡	.18‡	.04	−.08	.36‡	.29‡	.25‡

*$p < .05$. †$p < .01$. ‡$p < .0001$

ticularly large. Our 55.02% male representation differed from the U.S. Census representation of 48.89% by 6.13%. Our 84.22% Caucasian representation differed from the U.S. Census representation of 71.37% by 12.85%. It appears that gender parity was more nearly realized than racial parity. The demographics of the Internet are expected to more closely approximate that of the U.S. Census given that the number of Internet users is approximately doubling each year. The Internet population is closer in age, is more racially diverse, and is at least as balanced on gender as college populations are (Gartner Group, 2000; U. S. Census, 2000). Even if the demographics for all Internet users remain biased, the number of Internet users is so large that a subsample with the desired demographics can likely be found, as was done in the present study. An important implication of these results is that investigators can use the Internet to access more diverse and representative populations of research participants (cf. Jerome et al., 2000).

Multicultural psychology arose primarily because mainstream psychology emphasized internal validity more than it did external validity. Most investigators have not demonstrated that their findings generalize over important dimensions of diversity (Sue, 1999). Although data from mainly white college sophomores provide some empirical support for the use of various psychological tests and treatments, these data do not validate them generally. Validation requires evidence of generalizability, and that requires access to a more diverse population of research participants than has previously been available to most investigators. The Internet increases access to such participants.

Statistical power is directly proportional to sample size. The present study's ability to generate 2,230 complete data sets in less than a year indicates that it is possible to collect data from large samples in relatively short time periods by using the Internet. This number compares favorably with that for well-financed research projects. A final contribution of the present study is that it constitutes a positive replication of Watt and Ewing's (1996) data.

REFERENCES

Bailey, R. D., Foote, W. E., & Throckmorton, B. (2000). Human sexual behavior: A comparison of college and Internet surveys. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 141-168). San Diego: Academic Press.

Batinic, B., Reips, U.-D., & Bosnjak, M. (Eds.) (2002). *Online social sciences*. Cambridge, MA: Hogrefe & Huber.

Birnbaum, M. H. (Ed.) (2000a). *Introduction to behavioral research on the Internet*. Upper Saddle River, NJ: Prentice-Hall.

Birnbaum, M. H. (2000b). *Psychological experiments on the Internet*. San Diego: Academic Press.

Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin*, **114**, 533-541.

Buchanan, T. (2000). Internet research: Self-monitoring and judgments of attractiveness. *Behavior Research Methods, Instruments, & Computers*, **32**, 521-527.

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

Buchanan, T., & Smith, J. L. (1999b). Using the Internet for psycho-

logical research: Personality testing on the World-Wide Web. *British Journal of Psychology*, **90**, 125-144.

CARLTON, L. G., CHOW, J. W., EKKEKAKIS, P., SHIM, J., ICHIYAMA, R., & CARLTON, M. J. (1999). A Web-based digitized video image system for the study of motor coordination. *Behavior Research Methods, Instruments, & Computers*, **31**, 57-62.

CRONK, B. C., & WEST, J. L. (2002). Personality research on the Internet: A comparison of Web-based and traditional instruments in take-home and in-class settings. *Behavior Research Methods, Instruments, & Computers*, **34**, 177-180.

DAVIS, R. N. (1999). Web-based administration of a personality questionnaire: Comparison with traditional methods. *Behavior Research Methods, Instruments, & Computers*, **31**, 572-577.

DIENER, E., EMMONS, R. A., LARSEN, R. J., & GRIFFIN, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, **49**, 71-75.

FARMER, R., & SUNDBERG, N. D. (1986). Boredom proneness—The development and correlates of a new scale. *Journal of Personality Assessment*, **50**, 4-17.

GARTNER GROUP (2000, October 30). *Gartner says average U.S. Internet user is 41 years old with an income of $65,000* [Press release]. Stamford, CT: Author. Retrieved January 15, 2001, from http://www.gartner.com/5_about/press_room/pr20001030a.html.

GHISELLI, E. E. (1964). *Theory of psychological measurement*. New York: McGraw-Hill.

GHISELLI, E. E., CAMPBELL, J. P., & ZEDECK, S. (1981). *Measurement theory for the behavioral sciences*. New York: W. H. Freeman.

GORDON, M. S., & ROSENBLUM, L. D. (2001). Audiovisual Speech Weblab: An Internet teaching and research laboratory. *Behavior Research Methods, Instruments, & Computers*, **33**, 267-269.

HONAKER, L. M. (1988). The equivalency of computerized and conventional MMPI administration: A critical review. *Clinical Psychology Review*, **8**, 561-577.

HUDSON, W. W., HARRISON, D. F., & CROSSCUP, P. C. (1981). A short-form scale to measure sexual discord in dyadic relationships. *Journal of Sex Research*, **17**, 157-174.

JEROME, L. W., DELEON, P. H., JAMES, L. C., FOLEN, R., EARLES, J., & GEDNEY, J. J. (2000). The coming of age of telecommunications in psychological research and practice. *American Psychologist*, **55**, 407-421.

JÖRESKOG, K. G., & SÖRBOM, D. (1996). *LISREL 8 user's reference guide*. Chicago: Scientific Software International.

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.

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: Academic Press.

NOLEN-HOEKSEMA, S., & MORROW, J. (1991). A prospective study of depression and posttraumatic stress symptoms after a natural disaster: The 1989 Loma Prieta earthquake. *Journal of Personality & Social Psychology*, **61**, 115-121.

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.

REIPS, U.-D. (2001). The Web Experimental Psychology Lab: Five years of data collection on the Internet. *Behavior Research Methods, Instruments, & Computers*, **33**, 201-211.

SCHMIDT, W. C. (1997). World-Wide Web survey research made easy with WWW Survey Assistant. *Behavior Research Methods, Instruments, & Computers*, **29**, 303-304.

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**, 496-505.

SNELL, W. E., & PAPINI, D. R. (1989). The sexuality scale: An instrument to measure sexual-esteem, sexual-depression, and sexual-preoccupation. *Journal of Sex Research*, **26**, 256-263.

SNYDER, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality & Social Psychology*, **30**, 526-537.

SNYDER, M., BERSCHEID, E., & GLICK, P. (1985). Focusing on the exterior and the interior: Two investigations of the initiation of personal relationships. *Journal of Personality & Social Psychology*, **48**, 1427-1439.

SNYDER, M., & GANGESTAD, S. W. (1986). On the nature of self-monitoring: Matters of assessment, matters of validity. *Journal of Personality & Social Psychology*, **51**, 125-139.

STANTON, J. M. (1998). An empirical assessment of data collection using the Internet. *Personnel Psychology*, **51**, 709-725.

SUE, S. (1999). Science, ethnicity and bias: Where have we gone wrong? *American Psychologist*, **54**, 1070-1077.

U.S. CENSUS BUREAU (2000, November 2). *U. S. Census Bureau: College enrollment of students 14 to 34 years old, by type of college, attendance status, age, and gender: October 1970 to 1998* [Report]. Washington, DC: Author. Retrieved January 22, 2001, from the World-Wide Web.

WATT, J. D., & EWING, J. E. (1996). Toward the development and validation of a measure of sexual boredom. *Journal of Sex Research*, **33**, 57-66.

ZUCKERMAN, M. (1979). *Sensation seeking: Beyond the optimal level of arousal*. Hillsdale, NJ: Erlbaum.

(Manuscript received April 4, 2002;
revision accepted for publication March 30, 2003.)