Research on the Internet:
Validation of a World-Wide Web
mediated personality scale

TOM BUCHANAN and JOHN L. SMITH
University of Sunderland, Sunderland, England

Two studies were performed to assess the validity of a World-Wide Web (WWW) measure of self-monitoring. In Study 1, Usenet Newsgroups likely to be read by high and low self-monitors were identified and a comparison was made of the extent to which contributors engaged in a form of self-presentation (use of *handles* or *screen names*) likely to be influenced by self-monitoring tendencies. Handles were used significantly more frequently in the *high self-monitoring* Newsgroups, supporting the distinction made. In Study 2, participants recruited through these sets of Newsgroups completed the WWW-mediated test. Those from the high self-monitoring groups scored significantly higher. Self-reports of self-monitoring behavior also reflected scores on the scale. The results are interpreted as demonstrating the construct validity of the instrument used and the viability of criterion-group-oriented methods in Internet-mediated research.

There is growing interest in the use of the Internet as a medium through which to conduct various forms of psychological research (e.g., Buchanan & Smith, 1999; Hewson, Laurent, & Vogel, 1996; Krantz, Ballard, & Scher, 1997; Pasveer & Ellard, 1998; Schmidt, 1997; Smith & Leigh, 1997; Szabo & Frenkl, 1996). Reasons for this include the fact that the Internet offers access to very large numbers of potential participants at a very low cost to the experimenter, that experiments may be automatically run on line, and that individuals with particular characteristics of interest to researchers may be easily located (see, e.g., Hewson et al., 1996; Smith & Leigh, 1997; Szabo & Frenkl, 1996). Schmidt (1997) provides an up-to-date description of how such research may be done in practice. However, although it is clear that the World-Wide Web (WWW) and associated Internet technologies (e.g., e-mail, Usenet, IRC, virtual environments) do present exciting possibilities, it is also clear that there are many potential problems for Internet-mediated research, which must be addressed before it can become part of the empirical toolbox of mainstream psychological research.

A key concern is the validity of the instruments used for on-line research. Questionnaire-based instruments, such as surveys and personality scales, translate remarkably well to the new medium and are seeing wide use. A traditional paper-and-pencil questionnaire can easily be reproduced as a simple form on a WWW page. Respondents, recruited through various means and usually in locations remote from the experimenter, fill out the form using their browser software. Their responses are then transmitted to the experimenter and may be automatically processed in a variety of ways (see, e.g., Batinic, 1997).

However, the validity of such instruments can be compromised by a number of factors. These include the nature of the likely sample (which will vary, depending on recruiting strategy, but is unlikely to be representative of the general population and is likely to incorporate a number of known and unknown biases), the volunteer status (Hewson et al., 1996) and motivation of the participants (Oakes, 1972, has shown that volunteer and coerced samples may produce different findings), situational factors associated with the experimental environment (Reips, 1996, and Buchanan & Smith, 1999, discuss this at greater length), and the problem of repeat participation or mischievous responding (Schmidt, 1997). Each and any of these factors can have an impact on the power of a psychometric test to reliably and validly measure the construct it purports to address. Prior to using an on-line test in Internet-mediated research, one must, therefore, establish that it is sufficiently robust to such challenges that tests of the relevant hypotheses are not compromised.

Thus, the work reported in this paper is motivated by the need to establish the equivalence of on-line personality scales and their conventional antecedents (see Smith & Leigh, 1997, for an example of the validation of a survey instrument). One cannot simply assume that, when a traditional paper-and-pencil test is implemented through the Internet, the conventional and on-line versions will be equivalent. Work on the translation of tests for use on stand-alone computers has shown that such an assumption is not always valid. Cronbach (1990) concludes that "the conventional and computer versions of a test do usually measure the same variables, but difficulty or reliabil-

ity can easily change. Whether the computer version is 'the same test' must be queried with each instrument in turn psychologically" (p. 48). For the reasons mentioned above, this comment is particularly pertinent to cases in which the computerized test is administered via the Internet.

Self-Monitoring

One instrument that has been adapted for administration via the Internet is the revised Self-Monitoring Scale (SMS-R; Gangestad & Snyder, 1985). The SMS-R is a measure of the tendency to observe and modify expressive behaviors and self-presentation. Individuals high in self-monitoring are typically sensitive to social and situational cues and adjust their behavior accordingly. Low self-monitors, on the other hand, lack either the ability or the motivation to do this and tend to behave in ways consistent with their stable personality attributes or internal states (Snyder & Gangestad, 1986). The scale has been widely used and has been found to be related to a variety of social behaviors, including mate selection (Snyder, Berscheid, & Glick, 1985), behavior in romantic relationships (Snyder & Simpson, 1984), and career choice (Snyder & Gangestad, 1982).

We would also expect self-monitoring to influence certain aspects of social behavior on the Internet. A defining characteristic of the high self-monitor is that they are likely to engage in expressive self-presentation. They will tend to portray themselves in ways consistent with images they wish to adopt, in order to manipulate others' perceptions of them. This is something that many Internet users characteristically do, through the use of *screen names* or *handles*.

Goffman (1959) defines *front* as "that part of the individual's performance which regularly functions in a general and fixed fashion to define the situation for those who observe the performance" (p. 32). In order to distinguish the more individualistic aspects of front from the setting in general (which would include furniture, decor, and the like), he uses the term *personal front*. This refers to the other items of expressive equipment that most intimately identify the performer him/herself. Goffman states that, "as part of personal front we may include: insignia of office or rank; clothing, sex, age, and racial characteristics; size and looks; posture; speech patterns; facial expressions, bodily gestures and the like" (p. 34). Although Internet handles did not exist in 1959, they would not be out of place with the other items on this list, were it to be updated for the modern age of electronic communication.

Because, even in ordinary interpersonal communication, there is seldom scope to check out fully the front presented in a particular interaction, it is often possible to manipulate the specific persona that is put forward. In normal life, there are limits to the extent to which this may be done, and a deterrent to bending the facts of the matter too far will be the loss of face upon disclosure, as well as the possibility of real social or legal sanctions (e.g.,

think of someone passing as a doctor or a member of the opposite sex when he or she fails to satisfy the appropriate conditions of role occupancy). However, although there may be limits on the extent to which extreme instances of impression management may take place, there will still be plenty of scope for the individual to place a specific spin on his or her basic persona through the manipulation of personal front.

Adoption of alternative or modified personae has historically been an important aspect of behavior on line, where there are fewer constraints on how one portrays oneself. Often, when interacting with (passively reading a message from, or actively "chatting" with) another person in an Internet environment, the only information we have about him or her is the name he or she chooses to go by. In Usenet Newsgroups (on-line discussion groups; see Allie, 1995), as well as other areas of the Internet, there is a tradition of using self-chosen nicknames (known as *handles* or *screen names*) rather than one's real name (although not all people choose to use such a nickname). Such handles often bear little resemblance to real names. They may be fantastic in nature or bear some meaning for either the user or others. In some multi-user virtual environments, such as fantasy game MUDs, the focus is usually on adopting a fictional persona and playing out that role. Schiano (1997) suggests that roles played in a more *social* virtual environment (LambdaMOO) are often "slightly idealized or fanciful versions of one's view of oneself" (p. 272). Something similar to this occurs in IRC and other Chat Rooms, where a nickname is chosen by the user as the front that is seen by those with whom he or she interacts: In such a situation, the focus is less on role-playing another character and more on choosing how other people see you (Bechar-Israeli, 1995, discusses the use of nicknames on IRC). Within the context of Usenet Newsgroups, the use of a handle serves the same function.

A Usenet handle may be construed as a device for keying into a persona that, to some degree, differs from that which might be suggested by the person's real name. With regard to nicknames, Harré (1983) suggests that, when they are given to a person, they illustrate a social norm. So, *Piggy* or *String Bean* identify bodily forms that are generally unacceptable. Nicknames are seldom self-selected, and they are often pejorative. Usenet handles differ from this insofar as they are chosen by the people themselves and may, therefore, take on a positive cast, in order to enhance the electronic presentation of self (or perhaps it would be better to say, the presentation of electronic self).

The link between self-monitoring and impression management has already been set out in some detail by Snyder (see Snyder, 1987, especially chap. 1). In a nutshell, high self-monitors are more likely to seize the opportunity to manage their impressions through the exploration of novel personal fronts than are low self-monitors. Insofar as Internet handles facilitate this, it will be high

self-monitors who will tend to deploy them in their electronic communications. This is a point we will return to.

The Self-Monitoring Scale (Revised)

The SMS-R is an 18-item forced-choice questionnaire based on the original Self-Monitoring Scale developed by Snyder (1974). It has satisfactory reliability, and there is considerable evidence for its construct validity: "A large number of empirical investigations have provided support for a wide-ranging set of hypotheses about the involvement of self-monitoring in diverse domains of cognitive, behavioral and interpersonal functioning" (Snyder & Gangestad, 1986, p. 134).

Buchanan and Smith (1999) compared the psychometric properties of Internet-mediated and paper-and-pencil versions of the test and found them to be equivalent: The two forms were closely matched in terms of factor structure, reliability, and other points of comparison. Although the findings of that study indicate that the Internet-mediated version has adequate reliability and the expected factor structure, they tell us little about the most important question: Does the instrument actually differentiate between high and low self-monitors? Is it a valid measure of this personality trait?

VALIDATION STRATEGY

It has been suggested that one of the major boons of using the Internet for research is that it makes it easy to identify and contact groups of people with particular characteristics of interest (Hewson et al., 1996; Smith & Leigh 1997) who might be rare in the general population. This suggests that a criterion group approach to validation might be appropriate. If groups of high and low self-monitors can be recruited, comparison of the scores they obtain on the instrument should tell us something about its validity. The success of this strategy, of course, hinges on whether the criterion groups actually do differ in self-monitoring tendency: This must be established through some means other than the instrument being validated. One way to do this is to identify groups of people whose on-line behavior is consistent with that expected of high or low self-monitors (for instance, in the use of handles).

An alternative approach to validation would be to examine whether scores on the instrument correlate in the expected manner with other indices of self-monitoring one might obtain from participants. Unfortunately, the absence of validated on-line measures of constructs associated with self-monitoring rules out the use of additional psychometric instruments. Some of the potential differences (e.g., the nature of the sample and the conditions under which the test is completed) between WWW and traditional tests and the problems associated with testing the WWW participants under laboratory conditions also militate against concurrent validation with a paper-and-pencil version of the test (Buchanan & Smith, 1999).

However, it is possible to ask participants about the extent to which they engage in activities that should be influenced by their self-monitoring tendencies (such as modification of behavior in response to situational cues; high self-monitors should be more likely to do this). These approaches are not mutually exclusive; both may be implemented within the same study. It was, therefore, decided that both should be simultaneously employed in examining the validity of the Internet version of the SMS-R.

Selection of Groups

As was mentioned above, it is important to identify groups of people who should be high or low in self-monitoring tendency, in order that their scores might be compared. Thus, recruitment notices may be posted in areas likely to be read by members of such groups.

There is a problem with this approach, in that any two groups of people may well differ on a number of dimensions in addition to self-monitoring. Thus, any observed differences between the two groups might be due to such other factors. However, comparing multiple groups of people who should be high self-monitors with multiple groups of people who should be low self-monitors will reduce this possibility and will maximize the likelihood that the major dimension on which the groups systematically differ is, in fact, self-monitoring. Therefore, we decided that it would be appropriate to use at least two criterion groups that comprise high self-monitors and two criterion groups comprising low self-monitors.

To this end, we examined the self-monitoring literature for examples of such groups. Examples of people who tend to have self-monitoring scores that are high, relative to the rest of the population, are actors (Snyder, 1974) and obese people (Younger & Pliner, 1976). Although examples of people characteristically low in self-monitoring are harder to find, shy people tend, as a group, to score lower than the nonshy (Pilkonis, 1977). The only other group we are aware of that has empirically been shown to comprise low self-monitors is hospitalized psychiatric patients (Snyder, 1974). These seemed to be an unsuitable group for use in this study, not least because hospitalization would necessarily limit their opportunity to participate.

Thus, the second group of low self-monitors was identified on conceptual, rather than empirical, grounds. Snyder (1987) has supplemented the empirical findings with conceptual reasons why the groups mentioned above should score as they do. The prototypical low self-monitor is "less attentive to social information about situationally appropriate self-presentation . . . expressive self presentations seem, in a functional sense, to be controlled by inner attitudes, dispositions, and values, rather than to be molded and shaped to fit the situation" (p. 14). This definition suggests that low self-monitors are likely to express (and behave in ways consistent with) their attitudes and ideologies (irrespective of what those attitudes

are). A low self-monitor who believes in an ideal will speak out in support of that ideal and behave in ways consistent with it, rather than "going with the flow" of society. Snyder (1987) suggests that "low self-monitors excel in the domain of correspondence between private attitudes and public behaviors" (p. 39). People with strong commitments to single-issue politics would fall into this category. We decided that people consistently expressing strongly held beliefs would be likely to be low self-monitors, and we chose environmental campaigners as being an example of such a group.

The criterion groups in our high self-monitoring condition were thus obese people (Group A) and actors (Group B). The groups in the low self-monitoring condition were shy people (Group C) and people with a strong interest in the environment (Group D). We then identified Usenet Newsgroups likely to be read by members of these groups. The Newsgroups alt.support.big-folks and soc.support. fat-acceptance were deemed likely to be read by people falling into Group A. The Newsgroups rec.arts.theatre. misc and rec.arts.theatre.plays were deemed likely to be read by Group B. The readership of alt.support.shyness was considered likely to include a large number of people falling into Group C, whereas uk.environment and talk.environment were deemed likely to be read by Group D.

Establishing Differences
Between Criterion Groups

On conceptual (and, in most cases, empirical) grounds, we would predict differences in the likely self-monitoring tendencies of readers of the selected Newsgroups. However, if they are to be used as criterion groups in a validation study, it is important to establish that this is actually the case. Given that expressive self-presentation is a major element of self-monitoring, it seemed that examination of how people presented themselves in these Newsgroups might bear fruit.

We have already mentioned the use of handles or screen names in Usenet Newsgroups and noted that such behavior is likely to be influenced by self-monitoring tendencies. Although many contributors to Usenet Newsgroups use such screen names, not all choose to do so. If use of a handle is an example of expressive self-presentation, someone high in self-monitoring tendency would be more likely to use a handle, instead of his or her real name, than would a low self-monitor.

Therefore, one would expect more use of handles in a Newsgroup predominantly read by high self-monitors than in one read by low self-monitors. This suggests that the appropriateness of the Newsgroups listed above as sources of participants for our high and low self-monitoring criterion groups may be assessed by examining the relative frequency of handles and real names in each Newsgroup.

STUDY 1

Method

Participants. Participants in this study were 415 people who had posted messages to the Newsgroups listed above that were visible

(had not expired) on the date of our survey. Nothing is known about the participants, other than the name associated with the message.

Procedure. A "snapshot" of the contributors to the Newsgroups listed above was taken as follows. On November 10, 1997, we examined all the (noncommercial) articles available on the local Usenet news server[1] that had been posted to the Newsgroups in question. The author of each was noted, and a list of unique contributors was thus generated for each Newsgroup. The identifier (the name that appeared in the "From" line of the message header) of each contributor was then coded by two independent raters blind to the groups from which they had been drawn. The identifiers were classified as real names, as handles (according to the definition given above), or as being ambiguous. Where the identifiers were single forenames (e.g., Tom or John), they were classified as ambiguous, pending further investigation.

In this initial coding, 322 of the 368 unique identifiers were assigned to the same category by both raters (87.5% agreement). Most cases in which differences arose were cultural references known to one rater, but not to the other (e.g., names of fictional characters from films). All the identifiers classified by both raters as ambiguous (31, all being single forenames) or on which the raters disagreed (46) were then investigated further by examination of message headers and signatures in the postings. In 22 cases sufficient information was found to assign the identifier to one group or another. Those identifiers that could not be so resolved were classified as real names, as were identifiers like "Default Mail User," automatically generated by Internet software. This conservative approach was adopted because we had no evidence that these people were engaging in expressive self-presentation.

Results

In the high self-monitoring (Groups A and B) Newsgroups, a total of 71 handles and 164 apparent real names were seen; in the low self-monitoring Newsgroups (Groups C and D), 37 handles and 143 apparent real names were seen. A chi-square test for independence was carried out on the resulting 2 × 2 table (self-monitoring category × identifier type). This indicated that, for these Newsgroups, there was an association between self-monitoring category and frequency of use of handles: People posting messages in groups identified as likely to be read and contributed to by high self-monitors were more likely to use handles than were people posting to groups deemed likely to be used by low self-monitors [$\chi^2(1, N = 415) = 4.94, p < .027$].

Discussion

If the use of handles is deemed to be an example of expressive self-presentation—or self-monitoring—this finding supports the classification of Newsgroups made above and suggests that they are, indeed, suitable locations from which to recruit participants. These findings add weight to the (already conceptually strong) case for these as appropriate criterion groups. Notwithstanding any other potential differences between these groups, the weight of the evidence from previous empirical and theoretical work (e.g., Snyder, 1987) and from the findings of Study 1 suggests that members of these criterion groups really do differ in self-monitoring tendency in the manner suggested. Having established this, it is now possible to invite readers of these Newsgroups to complete the on-line version of the SMS-R and to use comparisons of their scores to draw inferences about its construct validity.

STUDY 2

Method

Materials. The instrument used in this study was a modified version of the Web-mediated version of the SMS-R developed by Buchanan and Smith (1999). This reproduces the items of Gangestad and Snyder's (1985) scale as an HTML[2] form that passes data to a CGI script[3] that processes and saves input, then provides feedback to the user (see Schmidt, 1997, for a general description of the principles involved in this).

Having read and affirmed that they understood an informed consent form, the participants were presented with the 18 items of the SMS-R and some demographic questions (sex, age, and whether they were students). Once the answers to these had been submitted, the participants were asked to provide (via a free text entry box) details of "any recent occasion(s) when you have modified your behavior in response to the situation." The participants were then presented with their scores and with sufficient information about self-monitoring to be able to interpret them. Some information about the purpose of the study was given, as was a full citation for Snyder (1987) for anyone interested in finding out more. This debriefing page also included three more free text entry boxes, in which the participants were invited to give responses to the following questions: "Do you agree or disagree with your score? Can you say why?", "Was there any particular reason you chose to take part in this study?", and "Do you have any other comments not addressed by the questions above?". Finally a message was printed, thanking the participants for their help and asking them not to do the test again.

Procedure. The participants were recruited by means of messages posted to the Usenet Newsgroups listed above. Each message (entitled "[Off Topic] Help Wanted, Please.") contained a brief invitation to participate in the study, the URL (Internet location) where the study could be found, and contact details for the experimenters. A different URL was used for each of the four groups, so that the group that each participant belonged to could be logged. This was done so that the class of Newsgroup in which a respondent had seen our posting (and, thus, the self-monitoring group to which they should be assigned for analytical purposes) was known. Messages were posted to all the groups on November 24, 1997. On December 4, 1997, owing to slow recruitment in that condition, messages were reposted to the Newsgroups in Group D.[4] On December 12, 1997, we reposted recruitment notices to the Newsgroups in Groups A, B, and D (sufficient participants had been recruited from Group C). Data collection continued until January 7, 1998, when we terminated the study.

Data screening and processing. Internet-mediated studies entail the risk that some of the observations are, for either accidental or mischievous reasons, actually multiple responses from the same people (see, e.g., Schmidt, 1997). For this reason, the unique Internet addresses of the computers used each time the test was completed were logged during the study, so that all but the first response coming from each address could be deleted. All duplicate and multiple responses were then deleted from the data file. Of the 224 responses obtained, 2 were thus discarded. Four further responses were also discarded; these came from people claiming, in the feedback on reasons for participation, not to be regular readers of the Newsgroups where they had seen the recruitment notice and from people claiming to have been e-mailed the URL by a friend. Thus, there remained 218 valid responses.

Participants. Of the 218 participants, 126 were male, and 92 were female. Seventy-five were students of some type. The participants ranged in age from 16 to 71 years (M = 32.42 years, SD = 10.77). Sixty-four were recruited via the Newsgroups in Group A, 51 from Group B, 76 from Group C, and 27 from Group D. There were, thus, 115 in the high self-monitoring condition and 103 in the low self-monitoring condition.

Examination of logged Internet addresses indicated that, although the participants came from a wide geographical range (countries identified included the United Kingdom, the Netherlands, Belgium, New Zealand, Australia, Sweden, Denmark, Austria, and Eire), the majority of accesses probably came from the United States.

Results

Comparison of self-monitoring scores across conditions. The mean self-monitoring scores of the participants in the high self-monitoring condition (Groups A and B) were compared with those in the low self-monitoring condition (Groups C and D) and were found to differ significantly [$t(216)$ = 3.79, p < .0005]. The participants recruited through Groups A and B scored higher (M = 10.02, SD = 3.61) than those recruited from Groups C and D (M = 8.08, SD = 3.95).

Self-monitoring score and reported self-monitoring behaviors. It was hypothesized that high self-monitors would be more able to report instances of behavior modification. Thus, the participants' responses to the request for examples of such instances were examined. Of the 218 participants, 114 provided some kind of response to this question.[5] These statements were examined independently by two judges and were coded as to whether they provided evidence of self-monitoring behavior (n = 99), whether they denied any self-monitoring behaviors (n = 7), or whether they were largely irrelevant (n = 8). A comparison of the codings of the two judges indicated 100% agreement.

It seemed that the majority of people were able to summon up an example of some kind (although some indicated that it was not typical behavior for them, suggesting that even a characteristically low self-monitor will modify his or her behavior upon occasion). However, a few denied ever engaging in behavior modification. Theoretically, the self-monitoring scores of such people should be very low, and this was found to be the case (M = 4.86, SD = 2.34). The scores of these participants differed significantly [$t(104)$ = 3.29, p < .001] from those who had supplied an example (M = 9.65, SD = 3.78). Since this finding is based on a small number of individuals, it should be interpreted with caution (although the proportion of people in the wider population with self-monitoring scores this low is similarly small).

Association of self-monitoring scores with other variables. Information as to the age and sex of the participants was also gathered. Self-monitoring tendencies are not believed to be meaningfully associated with either of these demographic variables. In the present sample, males and females did not differ in self-monitoring scores [$t(216)$ = 0.94, p > .34], and self-monitoring score was not significantly correlated with age [$r(216)$ = .066, p > .33].

Responses to other questions. The participants were given the opportunity to comment on the extent to which they agreed with their scores. Of the 116 who gave a response to this question, 91 said they agreed with their scores. Thirteen said that they disagreed, whereas the remainder either expressed no opinion or were unsure. One

hundred and seven participants responded to the query about why they had chosen to take part. The most common reason reported was curiosity (29% of the cases), the next most common being a liking for taking part in tests or surveys (15%). Fifty-six participants chose to make an additional comment. These included vacuous statements (9%), critiques of the research (16%), requests for information (11%), and salutations or expressions of good wishes to the experimenters (9%).

Discussion

The results of this study suggest that the instrument used is a valid measure of self-monitoring. Groups of people who should score differently on the instrument do seem to do so. Although most people seem capable of giving examples of behaviors associated with self-monitoring, the scores of the very small number of people who claim not to behave in such ways are very much lower than those of the rest of the sample. Correlations between self-monitoring scores and other variables are such as one would expect on the basis of the literature. Finally, the majority of the respondents indicated that they agreed with their scores. This could be interpreted either as evidence of face validity or as simply the well-known *Barnum effect*.

GENERAL DISCUSSION

Study 1 seems to demonstrate that location of criterion groups via Usenet is possible, whereas Study 2 seems to demonstrate the construct validity of the Internet-mediated version of the SMS-R. The evidence suggests that the instrument is robust to the challenges of conversion to this format and that it taps into the same construct as its paper-and-pencil antecedent.

However, Study 2 also demonstrates a potential problem with the purity of criterion groups recruited via Usenet. A small number of the participants indicated that they were not actually members of the population we wished to sample from. These were people who had stumbled upon the recruitment notice or had been sent the URL by a friend (this latter would seem, on the basis of anecdotal evidence, to be a not uncommon occurrence). Although the number of participants excluded from the study for this reason was small, this does suggest that any study using a criterion group approach should implement some kind of check to ensure the purity of the sample (such as asking people how they heard about the research).

Our experiences in this study also permit some other methodological observations to be made. Although the amount of traffic (i.e., the number of articles posted each day) may be a reasonable index of its size, the actual readership of a Newsgroup can only be estimated. This means that one cannot tell what proportion of the target population one has actually sampled (Batinic, 1997). We observed uneven responses across the Newsgroups we tar-

geted. It is likely that the number of responses one obtains will be influenced by a number of factors, including Newsgroup size, nature of readership, nature of the research topic and the instrument, and recruitment strategy. This is an area in which future research would be useful.

The sample sizes one may obtain through such a recruitment strategy may also be limited: It is likely that repeated postings to the same Newsgroups will produce fewer and fewer respondents and will risk diminishing the goodwill of readers in the same way that commercial "spam" postings do. Therefore, although the present findings suggest that Newsgroups may be good places to find and recruit from special populations, they also seem to have limitations. Internet researchers might be well advised to consider alternate sampling strategies, such as recruitment through Web site links (Stones & Perry, 1997, have demonstrated that this is also a viable way of sampling from special populations).

Another point is the problem of multiple responding by the same participants (see, e.g., Pasveer & Ellard, 1998; Schmidt, 1997). Although there was only a small number of such cases in the present study, we have seen other instances in which the rate has been much higher (e.g., Buchanan & Smith, 1999). It is possible that the likelihood of multiple responding may depend on the nature of the study or the population being sampled, but it seems clear that any Internet-mediated research must implement some kind of check or safeguard against it.

Buchanan and Smith (1999) demonstrated that an Internet-mediated version of a personality scale, the SMS-R, had equivalent psychometric properties to a traditional version. The present findings suggest that the instrument also seems to be a valid measure of the construct it purports to address. Further evidence for construct validity can only come from further use for psychological research in an Internet environment. The broader implication of this research is that a Web-mediated personality scale can, it seems, be robust to the challenges to validity we have noted. Although this is something that must still be ascertained for other instruments adapted for use on the Internet, the early signs are promising.

REFERENCES

ALLIE, D. A. (1995). The Internet and research: Explanation and resources. *Journal of Mind & Behavior*, **16**, 339-368.

BATINIC, B. (1997). How to make an Internet based survey? In W. Bandilla & F. Faulbaum (Eds.), *SoftStat '97 advances in statistical software 6* (pp. 125-132). Stuttgart: Lucius & Lucius.

BECHAR-ISRAELI, H. (1995). From (Bonehead) to (cLoNehEAd): Nicknames, play and identity on Internet relay chat. *Journal of Computer-Mediated Communication* [On-line serial], **1**(2). Available: http://jcmc.huji.ac.il/vol1/issue2/bechar.html

BUCHANAN, T., & SMITH, J. L. (1999). Using the Internet for psychological research: Personality testing on the World-Wide Web. *British Journal of Psychology*, **90**, 125-144.

CRONBACH, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: HarperCollins.

GANGESTAD, S. W., & SNYDER, M. (1985). "To carve nature at its joints": On the existence of discrete classes in personality. *Psychological Review*, **92**, 317-340.

GOFFMAN, E. (1959). *The presentation of self in everyday life*. Harmondsworth, U.K.: Penguin.

HARRÉ, R. (1983). *Personal being: A theory for individual psychology*. Oxford: Blackwell.

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.

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.

OAKES, W. (1972). External validity and the use of real people as subjects. *American Psychologist*, **27**, 959-962.

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.

PILKONIS, P. A. (1977). Shyness, public and private, and its relationship to other measures of social behavior. *Journal of Personality*, **45**, 585-595.

REIPS, U. (1996, October). *Experimenting in the World-Wide Web*. Paper presented at the 1996 Society for Computers in Psychology Conference, Chicago.

SCHIANO, D. J. (1997). Convergent methodologies in cyber-psychology: A case study. *Behavior Research Methods, Instruments, & Computers*, **29**, 270-273.

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

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

SNYDER, M. (1987). *Public appearances/private realities*. New York: Freeman.

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. (1982). Choosing social situations: Two investigations of self-monitoring processes. *Journal of Personality & Social Psychology*, **43**, 123-135.

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.

SNYDER, M., & SIMPSON, J. A. (1984). Self-monitoring and dating relationships. *Journal of Personality & Social Psychology*, **47**, 1281-1291.

STONES, A., & PERRY, D. (1997). Survey questionnaire data on panic attacks gathered using the World Wide Web. *Depression & Anxiety*, **6**, 86-87.

SZABO, A., & FRENKL, R. (1996). Consideration of research on Internet: Guidelines and implications for human movement studies. *Clinical Kinesiology*, **50**, 58-65.

YOUNGER, J. C., & PLINER, P. (1976). Obese–normal differences in the self-monitoring of expressive behavior. *Journal of Research in Personality*, **10**, 112-115.

NOTES

1. The computer on which messages forwarded from all around the world were stored.

2. Hypertext mark-up language, the code in which Web pages are written.

3. A program located on the Web server that may be automatically run when people connect to the appropriate Web page.

4. At this point, a third environmentally related Newsgroup (sci.environment) was added to this list because of its high level of traffic. Although this Newsgroup was not included in Study 1, a count made on the day of posting indicated that the ratio of handles (7) to real names (61) was similar to that seen in talk.environment and uk.environment This suggested that readers of the three Newsgroups were similar in terms of their self-monitoring tendencies.

5. Why did the other 104 not do so? That they were unable to seems unlikely, given the similar response rate to the other questions asked. Perhaps the cause was boredom or impatience to see their score.