

Online Personality Assessment: Equivalence of traditional and WWW personality measures.

*Tom Buchanan*¹

On-line personality tests are increasingly being used in a number of areas, from psychological research to commercial testing services, and interest in them seems to be growing. In addition to those who are professionally concerned with psychometric testing, people in general (or at least a sizeable number of Internet users) seem to be fascinated by the topic. There are newsgroups and other discussion forums devoted to personality and its assessment (many from a non-academic perspective), and numerous associated websites created by enthusiasts listing and linking to tests which one can do on-line.

A recent study (Buchanan, Goldberg & Johnson, 1999) recruited participants purely through entries in search engine databases (e.g. Altavista, Yahoo and Lycos): anyone who ended up participating must have actually been looking for personality tests. Data was acquired at the rate of around 20-25 responses a day. This clearly demonstrates that there are people "out there" interested in personality, and actively seeking tests they can take (an exploration of their reasons for doing this might be interesting).

Researchers, of psychological and other persuasions, obviously have much to gain: on-line tests give easy access to large samples and the opportunity to explore interesting questions. Occupational and "business" psychologists may well find many of their activities and working practices changed by this new medium for the delivery of psychometric testing services. Bartram (1998) has outlined one possible scenario, where tests are made available on-line and users would purchase their results from the publisher, and also a number of challenges the testing industry is likely to face. The end-users of psychometric testing services (employers, education, healthcare and other organizations) will also have an interest in this new way of obtaining the services they require. Effectively, anyone who might have had an interest in administering or taking a traditional pencil-and-paper personality test will be affected by on-line assessment.

For many of these groups, there are strong reasons to adopt on-line tests—accounts of such reasons are presented by, among others, Buchanan (in press); Buchanan & Smith (1999) and Smith & Leigh (1997). From the test

¹ Department of Psychology, University of Westminster, 309 Regent Street, London W1R 8AL, UK. Email tb99@iname.com

administrator's point of view, the mechanics of constructing on-line tests are relatively simple, if one has even a basic knowledge of HTML and some awareness of CGI scripting or Javascript. On-line resources and software to make it easy are available: this is no longer the preserve of the expert programmer, and the opportunity is there for anyone to create and use on-line tests.

Possible problems: Are the tests any good?

This may actually be a problem. As Bartram and Bayliss (1984) noted for traditional stand-alone computerized tests, it is probably easier to produce a slick-looking and interesting on-line test than it is to produce one which is valid and reliable. And this is happening. Many (probably most, and including the most widely used ones) of the instruments currently on-line lack any kind of credible psychological foundation or satisfactory psychometric properties. This is clearly not a desirable situation, for a number of practical, professional and ethical reasons.

One solution to this problem is to use established tests which we already know (from experience with pencil-and-paper versions) are reliable and valid. A number of researchers have adopted this strategy. However, there is still a question: do the psychometric properties of a traditional test transfer to an on-line version? There are reasons to suspect that this might not always be the case (Buchanan & Smith, 1999), especially in situations where the nature of the construct being tested might interact with the testing medium (e.g. computer anxiety). Research on off-line computerized tests has demonstrated that they are often (e.g. Bartram & Bayliss, 1984) but not always (e.g. Webster & Compeau, 1996) equivalent in terms of the constructs they measure.

As with other types of on-line research instrument (e.g. Krantz, Ballard & Scher, 1997; Krantz & Dalal, in press) it is therefore important to independently establish the psychometric adequacy of on-line tests (both in general and for specific instruments). Several researchers have done work which either bears on this issue or sets out directly to test it, and it is those efforts on which the current paper focuses.

Evidence of equivalence

Pasveer & Ellard (1998) probably published the first evaluation of an on-line personality scale. In developing a psychometric measure of "self-trust" they recruited and tested participants both through the Internet and traditional means (students filling out pencil and paper tests). They found that data gathered from the two types of sample were very similar, and that the internal consistencies and factor structures of the scales were comparable. The only important difference between the two forms was apparent greater variance in the WWW data (a finding replicated elsewhere).

Similarly, Davis (1998) compared web-mediated and traditional versions of two measures of negative affect. While WWW participants tended to report more negative

affect, satisfactory indices of reliability and indications of validity were found for both. In fact, if it is accepted that Internet users do tend to be more depressed (e.g. Kraut et al, 1998; Petrie & Gunn, 1998) the different levels of negative affect could be interpreted as evidence for the validity of these instruments. However, there are other possible explanations (e.g. increased self-disclosure).

Buchanan & Smith (e.g. 1999) have conducted a series of studies comparing an on-line version of Gangestad & Snyder's (1985) Self-Monitoring Scale (revised) with its paper-and-pencil antecedent. Buchanan & Smith (1999) found that the reliabilities and factor structures of the on- and off-line instruments were equivalent. They then went on to obtain some evidence for the validity of the on-line test by comparing the scores of people deemed likely to score high or low (Buchanan & Smith, in press) and replicating findings obtained by researchers using the traditional version (Buchanan, 1998).

Schwarzer, Mueller & Greenglass (1999) compared data from WWW respondents to a self-efficacy scale with responses from traditionally recruited and tested samples (German and Canadian students and teachers). The same construct seemed to be measured in each sample, whether on-line or off: similar psychometric properties and associations with other variables were observed.

Equivalence cannot be taken for granted

From these studies, it is tempting to conclude that on-line tests do measure the same constructs as their traditional equivalents, and do have satisfactory psychometric properties. However, the picture is not quite that simple, because evidence is starting to emerge that this is not always the case.

For example, Woolhouse & Myers (1999) compared data from web-based and traditional versions of a new personality test. Scale reliabilities were comparable and satisfactory. However, factor analyses indicated that while most items did load on the subscales they were meant to address, this was not always the case.

Similarly, recent work (Buchanan, et al, 1999) examining the psychometric properties of an on-line version of a five-factor personality inventory (Goldberg, 1999) found that the test's items usually did load most strongly on the factors they were intended to measure. However, in a number of cases items actually loaded more strongly on other factors. Reasons for this remain to be explored, but may well involve different interpretations of item content linked to language (or other) differences among respondents.

Encouragingly, however, preliminary analyses suggest that a revised version of the Goldberg scale does have acceptable psychometric properties and some degree of validity. A reasonable interpretation of the evidence to date would be that on-line and off-line tests can be, and often are, equivalent. However, equivalence is not something which we may take for granted and must be established for every instrument which is used.

Future directions

These are still early days for on-line personality testing. The potential is clearly there, but issues of validity do remain and there are many other questions which need to be addressed. For example, there is evidence that people's behavior on-line may be characterized as "disinhibited" (e.g. Joinson, 1998)—which may well affect the way they respond to certain types of questions. People may be more candid, and engage in greater self-disclosure when completing on-line tests. There are ethical questions associated with the use of on-line instruments. There are problems of self-selection, and sampling biases which still need further exploration.

So, there is still a lot of work to be done. But on-line tests seem to be here to stay. And while we should proceed with caution, we certainly can and should proceed.

References

- Bartram, D., & Bayliss, R. (1984). Automated testing: Past, present and future. *Journal of Occupational Psychology, 57*, 221-237.
- Bartram, D. (1998). *Distance Assessment: Psychological assessment through the Internet*. Paper presented at the 1998 British Psychological Society Division of Occupational Psychology Conference.
- Buchanan, T. (1998). *Internet Research: Self-monitoring and judgments of attractiveness*. Paper presented at the 1998 Society for Computers in Psychology Conference. Dallas, TX.
- Buchanan, T., Goldberg, L. R., & Johnson, J. A. (1999). *WWW personality assessment: Evaluation of an on-line Five Factor Inventory*. Manuscript in Preparation.
- Buchanan, T. (in press). Potential of the Internet for Personality Research. In M. H. Birnbaum (Ed.), *Psychological Experiments on the Internet*. Academic Press.
- 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.
- Buchanan, T., & Smith, J. L. (in press). Research on the Internet: Validation of a World-Wide Web mediated personality scale. *Behavior Research Methods, Instruments, & Computers*.
- Davis, R. N. (1998). *World Wide Web vs. paper-and-pencil administration of mood and personality questionnaires*. Manuscript submitted for publication.
- 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.

- Goldberg, L. R. (1999). *International Personality Item Pool: A Scientific Collaboratory for the Development of Advanced Measures of Personality and Other Individual Differences* [On-line]. Available: <http://ipip.ori.org/ipip/>
- Joinson, A. N. (1998). Causes and implications of disinhibited behavior on the Net. In J. Gackenbach (Ed.), *Psychology of the Internet* (43-60). New York: Academic Press.
- Krantz, J. H., & Dalal, R. (in press). Validity of Web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological Experiments on the Internet*. Academic Press.
- Krantz, J. H., Ballard, J., & Scher, J. (1997). Comparing the results of laboratory and World-Wide Web samples of the determinants of female attractiveness. *Behavior Research Methods, Instruments, & Computers*, 29, 264-269.
- Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukophadhyay, T., & Scherlis, W. (1998). Internet paradox: a social technology that reduces social involvement and psychological well-being? *American Psychologist*, 53, 1017-1031.
- 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.
- Petrie, H., & Gunn, D. (1998). *Internet "addiction": The effects of sex, age, depression and introversion*. Paper presented at the British Psychological Society London Conference, 15th December 1998.
- Schwarzer, R., Mueller, J. & Greenglass, E. (1999). Assessment of perceived self-efficacy on the Internet: Data collection in cyberspace. *Anxiety, Stress and Coping*, 12, 145-161.
- 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.
- Webster, J., & Compeau, D. (1996). Computer-assisted versus pencil-and-paper administration of questionnaires. *Behavior Research Methods, Instruments, & Computers*, 28, 567-576.
- Woolhouse, L., & Myers, S. (1999). *Factors affecting sample make-up: Results from an Internet-based personality questionnaire*. Paper presented at the 1999 British Psychological Society Social Psychology Section Conference.