

# Internet-based Questionnaire Assessment: Appropriate Use in Clinical Contexts

Tom Buchanan

*University of Westminster, London, UK*

---

**Abstract.** Use of internet-mediated questionnaire assessment is growing in behavioural telehealth provision. Literature on web-based psychological testing indicates that – while tests used online should be validated for online use – such measures can be reliable and valid. However, use of normative data (especially data from paper-and-pencil samples) presents problems. Studies have shown score distributions may differ across testing media, even when comparisons are made between equivalent samples. Data from 3 different projects illustrate such (substantive) differences, and the errors that would result from using established norms to interpret internet-mediated assessments. Given representativeness issues with “internet norms”, it is recommended that if online tests are used for clinical purposes, this should currently be done in a manner that does not rely on normative data. *Key words:* Internet, clinical, telehealth, psychometric, assessment, norms.

*Received October 21, 2002; Accepted February 10, 2003*

*Correspondence address:* Tom Buchanan, Department of Psychology, University of Westminster, 309 Regent Street, London W1B 2UW, UK. Tel: +44 207 911 5000 ext. 2017, Fax: +44 207 9115174, E-mail: buchant@wmin.ac.uk

---

The past several years have seen a proliferation of instruments purporting to measure psychological constructs become available via the internet. Some are clearly the work of amateur enthusiasts or are intended to be “just for fun”. Others are developed by professionals and are intended for more serious purposes. Some have clearly established psychometric properties. Others are unvalidated, conceptually confused, and may give entirely misleading feedback. It is not always easy to distinguish between these different types of test, and all of them can currently be found on websites related to mental health.

## What is online assessment?

A number of different ways have been suggested in which the internet and other technologies can be used in behavioural telehealth applications intended to bring mental health services to people for whom access would otherwise be difficult or impossible (Lazlo, Esterman, & Zabko, 1999; Nickelson, 1996). Various models of service provision have been suggested, ranging from situations where assessment and therapy are provided entirely online, to models where email, telephone or web-based services are used as an adjunct to face-to-face sessions. In many of these models, the ability remotely to administer and score questionnaire-based psychological tests (e.g. depression screening inventories) will be of value. For example, cognitive behaviour therapy practitioners could use online questionnaires in pre-treatment assessments of clients, then continue to communicate by email in subsequent work (e.g. monitoring completion of homework assignments).

Two main ways in which online tests can be used are diagnostically (as part of the assessment process) and to track changes in respondents over time (e.g. as outcome measures to evaluate success of therapy). The first – initial assessment – is probably the most prevalent (or at least, currently most visible) use of clinically-oriented questionnaires. Numerous self-help sites and entry-points to online therapy providers feature brief screening instruments which participants can complete online and receive feedback from. This feedback will usually give them their score and some advice based on it (e.g. if it is high, see a medical practitioner or therapist). Typically, there is some statement to the effect that this is not a formal diagnosis of a mental disorder.

While there might be concerns about how people react to feedback on psychological states or characteristics (Buchanan & Smith, 1999a; Buchanan 2002a) this model may well be of benefit to a number of people (this is clearly the view of the authors of such web pages). However, it becomes a source of potential concern when one considers issues such as those highlighted by Rabasca (2000), Griffiths (2001) or Buchanan (2002a): Who is scoring and interpreting the test? What happens to the data? Is the advice given affected by commercial interests (e.g. is the website sponsored by a pharmaceutical or managed-care company (Rabasca, 2000)). Does the test actually “work”, in terms of it being a reliable and valid index of the construct it is meant to measure? For the purpose of this article, I will focus on the last of these issues: Are online tests any good?

It is relatively easy to create a web page that displays the items of a psychological questionnaire in a browser such as Netscape or Internet Explorer. This is the case whether one is developing an entirely new instrument (e.g. Pasveer & Ellard, 1998) or simply implementing an existing paper-and-pencil test online (e.g. Buchanan & Smith, 1999a, Davis, 1999). It takes little more effort to enable instant scoring, feedback to test-takers, secure transmission of data and so forth. All of this is considerably less time-consuming than the process of test validation and ensuring fitness for purpose. And therein lies the rub: Are the tests currently being used online fit for the purposes they are being used for?

Several empirical papers have been published showing that online tests can be reliable and valid. In studies where comparisons with paper-and-pencil control conditions are made, or historical paper-and-pencil data are available, there are indications that online instruments usually do measure the same constructs as their traditional antecedents (e.g. Davis, 1999). In other cases, where no paper-and-pencil comparative data are available, there is evidence that online inventories behave in ways suggesting they have a degree of construct validity (e.g. Buchanan & Smith, 1999b; Buchanan, 2000a; Buchanan, Goldberg, & Johnson, 1999; Buchanan, Johnson, & Goldberg, 2003 – see Buchanan, 2001, or Buchanan, 2002a, for reviews).

While there is evidence that online tests can be reliable and valid, there is also evidence that psychometric properties may change subtly when a test is placed on the web (Buchanan, 2001, 2002a). For example, Buchanan, Johnson and Goldberg (2003), in an evaluation of a web-based version of a 5-factor personality inventory, found that a small number of the items loaded on the “wrong factors” (i.e. different factors to those they had loaded on in the offline development sample). In an evaluation of another web-based 5-factor inventory, Johnson (2000) found that a small number of the facet (subscale) level constructs appeared to load on dimensions other than those expected. Woolhouse and Myers (1999), comparing online and pencil-and-paper instruments of a Jungian personality inventory again found differences in the factor structures obtained for the 2 forms of the instrument. Buchanan, Ali, Heffernan, Ling, Parrott, Rodgers and Scholey (2002) examined the factor structure of an online implementation of a prospective memory questionnaire, and found that rather than the 4 factors predicted by previous research with the offline version of the instrument, only 2 could be clearly recovered.

There are also reasons to suspect that score distributions on computerized tests (and by implication online tests) may differ from their pencil-and-paper equivalents. Schulenberg and Yutzenka (1999) reviewed literature comparing computerized and paper-and-pencil adminis-

tration of measures of negative affect. Among the instruments used in the studies they discuss are the Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) and State-Trait Anxiety Inventory (Spielberger, Gorsuch, & Lushene, 1970) – both instruments with clear clinical applications. George, Lankford and Wilson (1992) administered these 2 instruments to a sample of college students in either computerized or pencil-and-paper formats, and found higher depression and trait anxiety scores in the computerized condition. This led them to recommend that separate norms be compiled for computerized versions of traditional tests.

The obvious conclusion from the evidence to date is that one cannot take the psychometric properties of an online test for granted, even if it is a direct translation of a valid conventional instrument. Even once acceptable psychometric properties are demonstrated, however, there are still issues that must be addressed.

### **Ways of using online tests**

There are 2 main ways in which assessments (whether online or offline) can be used to provide useful information about a person. The first (and most common, especially in contexts such as the use of psychometric tests for personnel selection, or screening for possible clinical symptoms) is nomothetically, across persons: the scores of individuals can be compared with normative data based on groups to which they belong. An example of such use would be administering an anxiety inventory to a person and comparing their score with norms or cutting scores to arrive at a judgement as to whether their anxiety level was sufficiently high to be considered abnormal and require therapeutic intervention. The second way is idiographically, within persons. An example of such use would be to monitor changes in a person's depression levels, perhaps comparing scores at the beginning and end of a course of treatment, and using degree of change as an outcome indicator.

In the context of online assessment, idiographic use of tests seems less problematic than nomothetic use. The reason for this is that nomothetic use of assessment instruments requires appropriate normative data. This is data that does not, at present, exist for the majority of tests a clinician might wish to use.

### **The problem of norms**

For many established and useful pencil-and-paper (and even computerized) instruments, considerable quantities of normative data exist. When a clinician administers a psychometric test as part of an assessment procedure, there is a ready table of scores they can use to inform their decision as to the existence and nature of a person's problem. So when a test is used on the internet, it might be tempting to use the same data to interpret scores obtained online. I am aware of instances where this has been done in the context of online research projects, and in all probability it is an increasingly frequent occurrence. However, there is considerable evidence that this is not appropriate: expected score distributions are likely to differ for online and offline samples.

Johnson (2000) compared the scores of 1766 adults who completed a 5-factor personality inventory via the internet with those of the 211 members of the community sample who completed the inventory in a paper and pencil format during its development (Goldberg, 1999). The internet sample scored significantly higher on Neuroticism and Openness, and lower in Agreeableness and Conscientiousness. The differences in this example could easily be attributed to differences between samples: the people on whom Goldberg's norms were based may well not have been representative of the population from which Johnson's respondents came. However, there are examples in the literature where differences between online and offline datasets cannot be so easily explained.

For example, Joinson (1999) randomly assigned college-student participants to complete questionnaires either via the internet or in a paper-and-pencil format. He found that respondents participating via the internet showed significantly lower levels of socially desirable responding (as evidenced by scores on a social desirability questionnaire). Similarly, Davis (1999) compared online and offline responses (from similar samples) to a measure of self-focused negative thought, and found higher levels (which one would associate with increased negative affect) in the sample tested via the internet. Fouladi, McCarthy, and Moller (2002) compared findings from 3 instruments measuring, respectively, attachment to father and to mother; perceived ability to reduce negative mood; and awareness of mood and mood regulation strategies. Participants, recruited from undergraduate psychology classes, were randomly assigned to either online or paper-and-pencil testing conditions. Analyses indicated that while the psychometric properties of the 2 versions were in general very similar, there were some small differences. Mean scores on 2 of the mood-related scales differed significantly, there were differences in means and variances on a number of items, and there were some distributional differences. Barak and Cohen (in press) investigated administration mode effects for a typology of career preferences. Participants (who were high school students randomly assigned to conditions) completed the measure on either 1 or 2 occasions, in either offline (paper-and-pencil) or online (via the internet) modes of administration. On all 6 of the inventory's scales, mean scores were higher in the online condition (these differences were statistically significant in 3 of the 6).

The studies described above lead one to suspect that differences in datasets from people tested online and offline may in some way be a function of the testing medium or mode of test administration. The implication is that online scores should not be compared with offline normative data. The potential consequences of making such comparisons can be inferred from the examples that follow.

### **Example 1: self-monitoring**

The personality trait of self-monitoring (Snyder, 1974; Snyder, 1987) is not one usually regarded as being of great relevance to psychopathology. However, it will serve as a useful initial example of some of the problems that use of inappropriate norms may present. Self-monitoring reflects enduring tendencies to perceive social cues from others and to moderate one's behaviour in response to them. High self-monitors typically pay considerable attention to such cues, and alter their behaviour and self-presentation accordingly. Low self-monitors, on the other hand, tend to behave much more in accord with their internal states or enduring traits and attitudes. The SMS-R has been implemented as an online instrument, and a series of studies have provided evidence supportive of the view that the online version is a valid measure of the self-monitoring construct (Buchanan, 2000a; Buchanan & Smith, 1999a, 1999b).

Snyder suggests that high self-monitors are identifiable as those whose scores fall in the upper quartile on the 18-item revision of the Self-Monitoring Scale (SMS-R; Snyder, 1987) with low self-monitors being those people scoring in the lower quartile. The cutting scores he uses to operationalize these classifications are 13 and 7 (based on traditional college student samples). However, these cutting scores do not accurately reflect the upper and lower quartiles of a combined sample of 1299 respondents (drawn from Buchanan & Smith, 1999a, and Buchanan, 2000a) who completed the SMS-R via the internet. For that sample, the top quartile was best approximated by scores of 12 and above (20.8%, 387 people, falling into this range) and the lower quartile was approximated by 6 and below (24.2%, 315 people, falling into this range). Only 270 had scores of 13 or over, while 432 had scores of 7 or below. Application of the cutting scores suggested by Snyder (1987) based on the traditional paper-and-pencil data would have led to misclassification of 234 people – 18% of the sample.

Potential explanations for the observed differences are not hard to come by. The proportion of

high and low self-monitors in the sample is likely to have been influenced by the recruiting strategies used (e.g. advertising for participants in internet discussion groups). The sample is thus highly unlikely to be one that is representative of the sample from which Snyder derived the suggested cutting scores. It is also possible that the self-presentation concerns characteristic of high self-monitors might influence the way they respond to questionnaires differentially in online and offline situations (Joinson, 1999, makes a similar suggestion with respect to this construct). Either, or both, of these possibilities could account for the differences described. In the absence of an explanation, however, one is still led to the conclusion of greatest importance for the central thesis of this paper: Snyder's offline cutting scores are unsuitable for use with data acquired online.

### **Example 2: everyday memory**

Sunderland, Harris and Baddeley (1984) published the Everyday Memory Questionnaire (EMQ), a 28-item instrument that measures self-reported everyday memory problems (e.g. forgetting someone's name, or where you have put something) with higher scores being indicative of more memory problems. It has been shown to discriminate between head-injured patients and controls (Sunderland et al., 1984). While there have been some challenges to the validity of self-report memory questionnaires (scores on self-report measures seldom correlate strongly with objective measures of memory performance obtained from neuropsychological tests) the EMQ has been used in numerous studies with clinical samples (e.g. closed head injuries; people with multiple sclerosis; stroke patients) and has clinical applications.

Cornish (2000) administered the EMQ to 277 undergraduate student participants, using traditional paper-and-pencil techniques, and reported a mean score of 90.2 ( $SD = 25.8$ ). Rodgers et al. (2003) administered the EMQ, along with other instruments, in a web-based study of the cognitive effects of recreational drug use. Two hundred and forty two of their 763 participants claimed never to have taken the recreational drugs cannabis and ecstasy (both of which were shown to have detrimental effects on aspects of self-reported memory performance). For these 242 drug-free participants, the mean score on the EMQ was 70.4 ( $SD = 26.2$ ). This mean score is substantially lower than that reported by Cornish. The difference (19.8 scale points) corresponds to an effect size (Glass's  $d$ ) of 0.76, which is "large" in Cohen's (1992) terms. Whatever the reason for the difference, it seems that use of norms based on Cornish's sample would be inappropriate in trying to interpret the online scores of Rodgers et al.'s sample.

### **Example 3: Hospital Anxiety and Depression Scale**

Preliminary data from a project currently in progress (Buchanan, 2002b) also demonstrate the potential problem of using inappropriate norms. Participants in this study completed an online version of Zigmond and Snaith's (1983) Hospital Anxiety and Depression Scale (HADS), along with the SMS-R and demographic questions.

The HADS is a brief measure with 7 questions each addressing anxiety and depression using a Likert style response scale. It is regarded as suitable for use in detection of clinically significant levels of dysphoria, and also for tracking changes in these conditions (Johnston, Wright, & Weinman, 1995). As such, it is an instrument remarkably suited to the types of suggested above for online tests. On each of the scales, scores from 8 to 10 are taken to indicate "possible clinical disorder" while scores from 11 to 21 (the maximum possible) are indicative of "probable clinical disorder" (Zigmond & Snaith, 1983; Johnson, Wright, & Weinman, 1995). Participants were recruited via links posted at the end of another online personality inventory ([www.personalitytest.org.uk](http://www.personalitytest.org.uk), which is indexed by search engines and linked to by a number of other sites) and on my home page. They were thus a self-selected sample with an interest in personality,

but there is no reason to believe they differed from a “normal” population in terms of the prevalence of anxiety or depression-related disorders.

For the purposes of the current analysis, the first 122 responses in the dataset were downloaded. To guard against multiple submissions from the same individuals, in 5 cases where there was more than 1 submission from the same IP number (internet address of the computer being used), all but the first submission were deleted (see Buchanan, 2000b; Schmidt, 1997, for discussion of this procedure). Following this, any respondents who indicated that their data should not be included in analysis were also deleted from the datafile. The remainder were scrutinised for unrealistic combinations of demographic data (e.g. very young people claiming to have doctoral degrees, people living in Antarctica). No such responses were found. The analysis that follows is thus based on the first 102 valid responses.

In this sample, 46.1% of respondents scored 8 or above and 13.7% scored 11 or above for anxiety, while 19.6% of respondents scored 8 or above and 4.9% scored 11 or above for depression. Mean scores were 7.47 (SD = 3.91) and 4.76 (SD = 3.51) for anxiety and depression respectively. In comparison, in the sample of cancer patients reported by Mooney et al. (1991), 27% scored 8 or above for anxiety, while 8.7% scored 8 or above for depression. Mean scores were 5.44 (SD = 4.07) for anxiety and 3.02 (SD = 2.98) for depression.

Comparing the 2 samples, the differences are consistent with medium effect sizes (in Cohen's 1992 terms) for both anxiety ( $d = 0.50$ ), and depression ( $d = 0.40$ ). It seems very unlikely that the current online respondents are really noticeably more anxious and depressed than Mooney et al.'s participants, who had just received a diagnosis of cancer. In addition, it seems unlikely that almost 14% of respondents in the current study are suffering from a clinically significant level of anxiety.

It may well be that the elevated anxiety and depression scores found in this sample are due to their demographic characteristics. It may also be that these people have an interest in psychological testing – and thus found the data collection site – because they are experiencing some degree of psychological distress. But the sheer proportion scoring above the possible and probable cut-off points for anxiety and depression suggests that scores on measures of negative affect really are inflated when they are completed via the internet, and furthermore that norms and cutting scores based on offline samples are simply inappropriate as a basis for judgements about the clinical significance of online scores.

A full evaluation of these data must await investigation of the psychometric properties of the online HADS, and in particular its factor structure (work by McCue, Martin, Buchanan, Rodgers, & Scholey, *in press*), suggests that the traditional 2-factor structure may not provide the best fit to the data, at least with certain groups of participants). The scales do however appear to have adequate reliability: in the present sample, values of Cronbach's alpha for the anxiety and depression scales were, respectively, 0.80 and 0.77. These are similar to the values reported by Andersson, Kaldö-Sandström, Ström, and Strömberg (in press), in a sample of people with tinnitus who completed the Swedish version of the HADS online. Andersson et al. also reported good test-retest reliability. Also consistent with the current data is Andersson et al.'s finding that their online participants scored higher on both anxiety and depression scales than did a comparison clinic sample. This led Andersson et al. to suggest that different cut-off points for internet administration of the HADS might be required, a suggestion reinforced by the present data.

### Possible reasons for differences

On the basis of the examples listed above, it would seem that there are numerous instances where mode of test administration affects expected score distributions. In some cases, the reasons may be mundane (sampling from different populations). In other cases, there may be explanations which are more psychologically interesting, and which parallel the reasons advanced by Schulenberg and Yutzenka (1999) for differences between (offline) computerized and traditional

tests. A strong candidate is increased self-disclosure. There is compelling evidence that people may disclose more about themselves when communicating via computers than face-to-face interactions (e.g. Weisband & Kiesler, 1996), a phenomenon which appears to extend to internet-mediated communication also (Joinson, 2002). This has actually been one of possible advantages suggested for online clinical work and has also led to the suggestion that online psychological questionnaires will actually give a better picture of the individual's real personality than traditional measures (Buchanan, Joinson, & Ali, 2002).

Another possibility is that of interactions between the construct of interest and the assessment medium. For example, it is probably not a good idea to assess computer anxiety using a computer.

A third possibility, advanced by Schulenberg and Yutrzenka (1999), is that increases in negative affect may be mediated by computer anxiety or aversion. This suggestion is supported by George, Lankford and Wilson's (1992) finding that levels of computer aversion were associated with depression scores in the computerized assessment conditions only, and by Tseng, Tiplady, Macleod and Wright's (1998) finding that computer anxiety can affect responses in different ways depending on the technology used to gather data from participants.

Whatever the reason(s), there is clearly a strong need for further research along the lines of studies conducted by, for example, Joinson (1999), Davis (1999), Cronk and West (2002) and Fouladi, McCarthy, and Moller (2002), in which matched participants are assigned to online and offline measurement conditions. This should be done for any construct one might wish to measure online, in a variety of populations, so that any mode-of-administration effects can be identified. As more constructs are investigated in this way, those assessment domains most likely to be affected by (or robust to) mode effects can be identified. This is important both from the perspective of understanding the causes of such effects, and from the perspective of facilitating decisions about the clinical significance of data obtained from individuals in online assessments.

## **Conclusions and recommendations**

It seems clear on the basis of the literature published to date that for any test used on the internet, in any way, reliability and validity need to be established. If the test is an adaptation of a traditional offline instrument, evidence that the offline version has satisfactory psychometric properties is not sufficient to allow one to assume they will apply to the online version as well. While most of the evidence to date does indicate that online adaptations of offline tests usually address the expected constructs, there have been sufficient indications of (usually small) differences (e.g. in factor structure, score distributions) to advocate caution, especially in instances where test use has real implications for people's well-being (c.f. Buchanan, 2002a).

Even when online tests appear to be psychometrically satisfactory, there are still issues that must be borne in mind. One of the most salient is the use of normative data. The literature cited, and new data presented, in this article very strongly indicate that established offline norms for established offline tests are not appropriate for use with online versions. In cases where one might wish to compare the scores of an individual to such norms, to aid in diagnosis, there is a very strong chance that the conclusion arrived at will be wrong. A simple web search will reveal many sites presenting screening inventories for various conditions (e.g. depression or anxiety). If normal offline population norms are used as the basis for these recommendations, many more people might be being advised that they may have a problem than is actually the case.

If one wishes to use tests nomothetically, comparing online scores to normative data, then normative data gathered online must be used. Even this, however, throws up difficult issues. Given world-wide access to the Internet – and thus potentially (local and national legislation permitting) to the services offered by an online therapist – it is unlikely that the concept of “general internet norms” is realistic, given the demographic, cultural and symptomatic diversity likely to be found in the “general internet population”.

A more appropriate approach might be gathering very large datasets, with detailed demographic information, so that individuals can be matched with reference groups tailored specifically for them (e.g. same age, culture, job, marital status, or any other variable which might affect the construct being measured). This is a suggestion that is entirely achievable with current technology and techniques – Buchanan, 2000a, is an example of it being done on a small scale. The difficulty is in gathering sufficient data to make it feasible. However, reports on very large datasets (e.g. Robins, Tracy, & Trzesniewski, 2001,  $n = 326,641$ ) are now beginning to appear in the literature, which is a promising sign for this approach.

Given the issues outlined above, a reasonable recommendation for “best practice” in the clinical use of online questionnaire-based tests might be that their use should currently be limited to idiographic (within-person) approaches that do not require comparisons with normative data. Examples of such approaches would be tracking changes on multiple administrations of the same test to a person, and instruments that are ipsative in nature (e.g. several popular measures of personality types or various sorts of preference, as commonly used in career counselling – c.f. Barak & Cohen, in press). Alternatively, practitioners who wish to conduct therapy online and also use norm-based assessments could adopt a mixed online/offline model – with initial assessments perhaps being conducted in person or by post, and subsequent client-therapist interactions occurring electronically.

Internet-based psychological questionnaires are not just a possibility: they are reality. There is considerable evidence that they can “work” but also evidence that psychometric properties need to be firmly established, especially if they are to be used in any way that might affect the well-being of individuals. There is also evidence that the conditions under which they are used must be considered. At this point in time, it seems prudent to advise against any clinical uses that rely on normative data, especially if those data were gathered using traditional paper and pencil techniques.

## References

- Andersson, G., Kaldö-Sandström, V., Ström, L. & Strömberg, T. (in press) Internet administration of the Hospital Anxiety and Depression Scale (HADS) in a sample of tinnitus patients. *Journal of Psychosomatic Research*.
- Barak, A. & Cohen, L. (in press) Empirical examination of an online version of the Self-Directed Search. *Journal of Career Assessment*.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561–571.
- Buchanan, T. (2000a). Internet research: self-monitoring and judgements of attractiveness. *Behavior Research Methods, Instruments, & Computers*, 32, 521–527.
- Buchanan, T. (2000b). Potential of the internet for personality research. In M. H. Birnbaum, (Ed.), *Psychological Experiments on the Internet*. (pp. 121–140). San Diego, CA: Academic Press.
- Buchanan, T. (2001). Online personality assessment. In U.-D. Reips & M. Bosnjak, (Eds), *Dimensions of Internet Science*. (pp. 57–74). Lengerich, Germany: Pabst Science Publishers.
- Buchanan, T. (2002a). Online assessment: desirable or dangerous? *Professional Psychology: Research and Practice*, 33, 148–154.
- Buchanan, T. (2002b). [Online HADS and SMS-R scores]. Unpublished raw data.
- Buchanan, T., Ali, T., Heffernan, T. M., Ling, J., Parrott, A., Rodgers, J. & Scholey, A. S. (2002, October). *Psychometric Properties of Online Self-report Memory Questionnaires: The EMQ and PMQ*. Poster session presented at German Online Research '02, Hohenheim, Germany.
- Buchanan, T., Goldberg, L. R. & Johnson, J. A. (1999, November). *WWW Personality Assessment: Evaluation of an Online Five Factor Inventory*. Paper presented at the meeting of the Society for Computers in Psychology, Los Angeles, CA.
- Buchanan, T., Johnson, J. A. & Goldberg, L. R. (2003). *Implementing a Five-Factor Personality Inventory for Use on the Internet*. Manuscript submitted for publication.
- Buchanan, T., Johnson, A. N. & Ali, T. (2002, October). *Development of a Behavioural Measure of Self-*



- Disclosure for Use in Online Research*. Paper presented at German Online Research '02, Hohenheim, Germany.
- Buchanan, T. & Smith, J. L. (1999a). 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. (1999b). Research on the internet: validation of a world-wide web mediated personality scale. *Behavior Research Methods, Instruments, & Computers*, 31, 565–571.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
- Cornish, I. M. (2000). Factor structure of the Everyday Memory Questionnaire. *British Journal of Psychology*, 91, 427–438.
- 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, Instrument, & 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.
- Fouladi, R. T., McCarthy, C. J. & Moller, N. P. (2002). Paper-and-pencil or online? Evaluating mode effects on measures of emotional functioning and attachment. *Assessment*, 9, 204–215.
- George, C. E., Lankford, J. S. & Wilson, S. E. (1992). The effects of computerized versus paper-and-pencil administration of negative affect. *Computers in Human Behavior*, 12, 159–166.
- Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. J. Deary, F. De Fruyt, & F. Ostendorf, (Eds), *Personality Psychology in Europe, Vol. 7*, 7–28. Tilburg, The Netherlands: Tilburg University Press.
- Griffiths, M. (2001). Online therapy: a cause for concern? *The Psychologist*, 14, 244–248.
- Johnson, J. A. (2000, March). *Web-based Personality Assessment*. Poster session presented at the 71st Annual Meeting of the Eastern Psychological Association, Baltimore, MD.
- Johnson, M., Wright, S. & Weinman, J. (1995). *Measures in Health Psychology: A User's Portfolio*. Windsor, UK: NFER-Nelson.
- Joinson, A. N. (1999). Anonymity, disinhibition and social desirability on the internet. *Behavior Research Methods, Instruments, and Computers*, 31, 433–438.
- Joinson, A. N. (2002). Understanding the Psychology of INTERNET behaviour: Virtual Worlds, Real Lives. Basingstoke, UK: Palgrave.
- Lazlo, J. V., Esterman, G. & Zabko, S. (1999). Therapy over the internet? Theory, research and finances. *CyberPsychology and Behavior*, 2, 293–307.
- McCue, P., Martin, C. R., Buchanan, T., Rodgers, J. & Scholey, A. B. (in press). *An Investigation into the Psychometric Properties of the Hospital Anxiety and Depression Scale in Individuals with Chronic Fatigue Syndrome*. *Psychology, Health and Medicine*.
- Moorey, S., Greer, S., Watson, M., Gorman, C., Rowden, L., Tunmore, R., Robertson, B. & Bliss, J. (1991). The factor structure and factor stability of the Hospital Anxiety and Depression Scale in patients with cancer. *British Journal of Psychiatry*, 158, 255–259.
- Nickelson, D. W. (1996). Behavioral telehealth: emerging practice, research and policy opportunities. *Behavioral Sciences and the Law*, 14, 443–47.
- 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.
- Rabasca, L. (2000). Self-help sites: a blessing or a bane? *APA Monitor on Psychology*, 31, 13.
- Robins, R. W., Tracy, J. L. & Trzesniewski, K. (2001). Personality correlates of self-esteem. *Journal of Research in Personality*, 5, 463–482.
- Rodgers, J., Buchanan, T., Scholey, A. B., Heffernan, T. M., Ling, J. & Parrott, A. C. (2003). *Patterns of Drug Use and the Influence of Gender on Self-reports of Memory Ability in Ecstasy Users: a Web-based Study*. Manuscript submitted for publication.
- Schmidt, W. C. (1997). World-wide web survey research: benefits, potential problems, and solutions. *Behavior Research Methods, Instruments, & Computers*, 29, 274–279.
- Sunderland, A., Harris, J. E. & Baddeley, A. D. (1983). Do laboratory tests predict everyday memory? A neuropsychological study. *Journal of Verbal Learning and Verbal Behaviour*, 22, 341–357.
- Schulenberg, S. E. & Yutzenka, B. A. (1999). The equivalence of computerized and paper-and-pencil psychological instruments: implications for measures of negative affect. *Behavior Research Methods, Instrument, & Computers*, 31, 315–321.
- Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, 30, 526–537.
- Snyder, M. (1987). *Public Appearances/Private Realities*. New York: WH Freeman.

- Spielberger, C. D., Gorsuch, R. L. & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Tseng, H. M., Tiplady, B., Macleod, H. A. & Wright, P. W. (1998). Computer anxiety: a comparison of pen-based personal digital assistants, conventional computer and paper assessment of mood and performance. *British Journal of Psychology*, 89, 599–610.
- Weisband, S. & Kiesler, S. (1996). Self Disclosure on computer forms: meta-analysis and implications. *Proceedings of the CHI '96 Conference on Human-Computer Interaction*, April 14–22, Vancouver.
- Woolhouse, L., Myers, S. (1999, September). *Factors Affecting Sample Make-up: Results from an Internet-based Personality Questionnaire*. Paper presented at the 1999 British Psychological Society Social Psychology Section Conference.
- Zigmond, A. S. & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, 67, 361–370.

Copyright of Cognitive Behaviour Therapy is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.