CHAPTER 1

Testing on the Internet: Issues, Challenges and Opportunities in the Field of Occupational Assessment1

Dave Bartram
SHL Group plc, UK

This chapter starts by considering what the Internet is and what it can offer in relation to testing and assessment in the work and organisational field. It then goes on to take a look into the future and consider a range of practical and good practice issues. In considering testing on the Internet we need to consider both the technical strengths and weaknesses of the Internet itself (as a transport medium) and the limitations that the WWW technology imposes on the design of tests and control over their delivery. Throughout the chapter, the emphasis will be on the use of computer-based and web-based testing in the field of occupational assessment. Other chapters in this volume consider applications in other fields, notably educational testing and testing for licensing and certification.

COMPUTER-BASED TESTING (CBT) BEFORE THE INTERNET

The main value of CBT historically has been in the area of report generation. Some of the earliest systems (back in the days before personal computers) were designed to automate the scoring and interpretation of instruments such as the MMPI. With the advent of the personal computer, we saw the development of computer-administered versions of paper and pencil tests. These provided some advantages over paper and pencil, in terms of control of administration, and some disadvantages (e.g. the need for sufficient hardware

1 Based on a keynote presentation to the ITC Winchester Conference, 13 June 2002.

Computer-Based Testing and the Internet: Issues and Advances.
to test groups of people). They also raised the question of equivalence with their paper and pencil counterparts. Most research (see Bartram, 2005; Mead & Drasgow, 1993) has tended to show that equivalence was not a major problem so long as the tests were not speeded measures of ability.

Bartram (1997) commented on the fact that, despite the potential offered by technology for novel forms of assessment, the literature on computer-based assessment (CBA) within occupational assessment settings has been largely confined to a small number of issues. These have been dominated by the issues relating to the parallel use of computer-based and paper-based versions of the same tests and use of computers to generate descriptive and interpretative reports of test results (for reviews, see Bartram and Bayliss, 1984; Bartram, 1987b, 1989, 1993, 1994, 2005).

**INNOVATION IN COMPUTER-BASED TESTING**

Despite the increasing sophistication of computer-based assessment systems, within the field of occupational assessment the tests they contain are, typically, computer implementations of old paper-and-pencil tests. Nevertheless, there has been innovation in the field and the consequences of that innovation are increasingly finding their way into commercial practice. Tests can be innovative in a number of different ways. The most obvious is where the actual test content is innovative. However, innovation can also occur in less obvious ways. The process used to construct the test may be innovative and rely on computer technology and the nature of the scoring of the items may be innovative. In practice there is an interaction between these different aspects of innovation, in that some of the most interesting developments in test content also involve innovation in how that content is created.

For computer-based testing, the most obvious examples of content innovation can be found where tests use sound or video to create multi-media items. Drasgow, Olson-Buchanan and Moberg (1999) describe a full-motion interactive video assessment, which uses video clips followed by multiple choice questions. Simulations can also be run on computer to provide realistic work-sample assessments. Hanson, Borman, Mogilka, Manning, and Hedge (1999), for example, describe the development of a computer-based performance measure for air traffic control selection and Bartram (Bartram & Dale, 1983; Bartram, 1987a) describes the use of a simplified landing simulator for use in pilot selection.

Innovation in content also relates to the use of more dynamic item types, for example, where drag-and-drop or other familiar Windows-based operations are used rather than the simple point-and-click simulation of paper-and-pencil multiple-response. A review of this area of innovation in item types is presented by Drasgow and Mattern in Chapter 3.

Innovation in content, however, is often associated with novel methods of content generation. Item generation techniques have provided the potential for a whole host of new item types as well as more efficient production of
conventional items. Bartram (2002a), and Goeters and Lorenz (2002) describe the use of generative techniques to develop a wide range of task-based and item-based tests for use in pilot selection. It is worth noting, however, that most of the developments in this area of innovation have occurred in areas where selection leads to very high cost training or into high risk occupations or both (as is the case for trainee military pilot selection). Innovation is expensive, and the sort of tests described in the papers referred to above have required extensive research and development programmes. However, as in all areas of testing, the lessons learned from this work will result in benefits in due course for the general field of occupational assessment.

Computer software provides for the recording of very detailed information about a test-taker’s performance. In addition to the response given to an item, we can record how long the person took to respond. We can also record information about choices made and changed during the process of responding. For more complex item types we can track the performance of the person as they work their way through a task or series of subtasks. Bartram (2002b) reported validation data from the use of a set of computer-based ability tests that were administered without any time limit. These were designed for use in a diagnostic mode for people entering further education training courses. Time to respond was normed independently for each item and response latency was scored together with accuracy to produce a measure of efficiency. This efficiency score had higher validity than the traditional number correct score.

While there has been some experimentation in the use of response latency data for checking response stability (Fekken & Jackson, 1988) and socially desirable responding (Holden & Fekken, 1988; George & Skinner, 1990a, 1990b), these approaches have not been developed into practical applications for general use in personnel selection or other areas of I/O assessment as yet.

Item Response Theory (IRT) has been with us since the 1980s (Lord, 1980), but its application has tended to be confined to educational and some large-scale occupational uses. It has not been generally applied in the area of occupational assessment until relatively recently. IRT has the considerable advantage of approaching test construction from the item level. Its application to routine occupational assessment has become possible with the advent of better data collection and data management procedures. IRT has many advantages over traditional methods; however it also comes with some costs: the need for larger samples of data with which to determine the properties of items. Although computer technology has provided the possibility of implementing adaptive testing using (in particular where it is based on IRT models) the impact of this on general test practice has been slight in the occupational field. The main reason why traditional classical fixed-item-set tests have held sway for so long has been one of infrastructure. Neither paper and pencil nor PC-based testing is well suited to adaptive testing and the use of large item banks. The Internet potentially changes all of this. There are clear signs that attitudes to CBA are changing as people come to appreciate the real benefits of technology for assessment, and as the technological infrastructure needed to support these applications becomes increasingly ubiquitous.
Development of Computer-Based Testing and Growth of the Internet

The use of computer-based testing is increasing rapidly. It has been helped not only by the development of better interfaces, but by the dramatic increases in volume of and accessibility to hardware. Access to the Internet is now available to you in your home for a few hundred pounds of capital outlay. In addition we have seen the advent of email and restricted Internet services on digital TV systems. The new millennium heralded the appearance of the first generation of WAP mobile phones, with their ability to access the Internet in a wireless environment and the promise of broadband 3G systems becoming available in the next few years. However, the pattern of development is not uniform around the world. Even where the technology is present, some users are more conservative than others in their adoption of that technology.

Computer networks have existed for a long time. The first use of a hyper-linked network by the US military occurred in 1957. Academic institutions in the UK joined in 1973 when University College London set up the first connection. The first commercial UK IP network was set up in 1989. At the start of the 1990s, Tim Berners Lee proposed the idea of using a standard graphical browser and a communication standard to provide access to data from any source, and so ‘invented’ the World-Wide Web (WWW). The Mosaic browser, the first of the WWW browsers, appeared in 1992. In 1994, Netscape was founded and a year later, Microsoft embraced the Internet, having previously dismissed it.

In many ways we can look on 1995 as the real beginning of widespread use of the Internet, the time at which it started to become part of the fabric of many people’s everyday lives. In the few years since then, the range of applications and volume of use have mushroomed. For all practical purposes, while the potential of the Internet has been known for many years, it has only just reached the stage of development at which that potential can begin to be realised. We are now at a significant watershed in its development for a number of reasons.

Within North America, Europe and Asia–Pacific, we now have widespread availability of inexpensive, high-powered computer systems. As the hardware has become more widespread, so the range of service providers has increased. Now it is as easy to get onto the net as it is to have a phone installed. Indeed, wherever a phone or a cable TV connection has been installed, an Internet connection can be made. Once on the net, you have access to information and services that were previously restricted to expert users or specialists. You can be your own travel agent; you can buy books and other goods from anywhere in the world; you can consult experts, read government reports, or find a new job.

The convergence towards common standards has made it commercially viable for service providers to offer users more and more sophisticated applications. The advances in technology have provided us with standard features we would hardly have dreamt of a few years ago: minimum screen
resolutions of 1024 × 768 and 32-bit colour resolution; real-time animation, video and sound capabilities; multi-tasking and so on.

Finally, we have also witnessed an increase in reliability. This is the key to the use of computers in testing. Though computer systems are still prone to crashes, hang-ups and network failures, we are moving rapidly closer to the point where the user expectation is that computers should operate reliably.

Tests and documents are essentially ‘soft’ products. As such they can be downloaded over the Internet to users. This means that the Internet can be used as a complete commercial solution for test publishers. There is no longer any need for printing and production, warehousing and postal delivery services.

More significant for testing, however, is the shift in locus of control provided by the Internet from the ‘client side’ to ‘server side’. For paper and pencil testing, publishers have had to provide users with test items, scoring keys and interpretation algorithms. As these are ‘public’, the danger of compromise and security breaches is high. Test users can (and do) pass these materials on to people who are not authorised to use them. All the test data also resides with the user. The process of developing norms, checking the performance of test items and carrying out validation studies is dependent upon costly procedures for recovering data from users. For the Internet this situation is reversed. The data and the intellectual property reside on the publisher’s server. The user has access only to those parts of the process that they need.

Time for a Revolution!

The infrastructure is now being built to support a radical change in the way testing is done. We are seeing the widespread availability and acceptance of computers; standardisation on operating systems and interfaces and the growth of the Internet as the generic communication medium.

Ipsos-Reid in their annual ‘Face of the Web’ survey (reviewed by Pastore, 2001c) actually claim that the revolution is over and that we are entering a ‘post-revolutionary’ phase as growth of the Internet market starts to level off in most of the developed regions around the world. Global Internet population was estimated at around 350 million at the end of 2000. For 2005 it is estimated that it will be around 500 million. The US still has more people online than any other country, but its share of global users is shrinking. Western Europe plus the remainder of the English speaking world (UK, Australia, Canada and South Africa) now rivals the US as a bloc. In 2000, 65% of the population in Sweden used the Internet in a 30-day period, while in Canada it was 60%. This surpassed the US (59%), for which the percentage remained constant from 1999 to 2000. However, in a poll conducted in 2002 (Ipsos, 2002), the percentage of US Americans going online in a 30-day period had jumped to 72%. Very rapid growth was also noted in Korea, urban China, urban India, France, Germany and the UK (where usage increased from 35% in 2000 to 50% in 2002).
Awareness of the Internet is now almost universal in North America, Australia, Europe and Japan. While awareness levels are low in urban areas of China, India and Russia, they are increasing rapidly. The 2002 Ipsos study, for example showed an increase from 9 to 19% in urban India usage (over a 30-day period) and from 21 to 30% in urban China between 2000 and 2002. Internet use in Asia–Pacific is expected to surpass that in the US by 2005 (Pastore, 2001d, quoting an International Data Corp. study). Asia–Pacific (excluding Japan) is forecast to exceed 240 million users from the 2001 base of 64 million, with a fundamental shift in the position of China.

While predictions are always dangerous, by 2010 I believe that in the field of occupational assessment we will see computer-managed and computer-delivered assessment becoming the default and paper-and-pencil testing the exception.

WHAT CAN THE INTERNET OFFER NOW?

In this section some current areas of application of the Internet are considered, and illustrated. The applications covered include:

- using the Internet to support test users
- assessment for development
- test practice and familiarisation
- recruitment and selection
- post-hire applications, including online 360-degree feedback and development.

Supporting Test Users

Most discussion of the Internet tends to focus around the delivery of tests to test takers online. However, the Internet also provides a new medium for distribution of test materials, reports and manuals, and for the automated collection of data. Even traditional paper and pencil materials can be delivered online as PDF format files using e-book publishing technologies.

Many publishers now use the Internet as a major source of support for test users. SHL, for example, delivers updates for its offline PC-based test systems over the Internet, together with norm updates and other technical data to support paper-and-pencil test users. As a medium, the Internet also provides a channel for users to communicate with each other (through user-groups) and with the publishers of materials. This communication is not just for traditional support purposes, but also provides the potential for supplier and client to have more of an ongoing dialogue through which the supplier can better target future research and development.
Self-Assessment and Practice

It is widely acknowledged (e.g. Kellett, 1991; Callan & Geary, 1994) that we should provide people who are to be assessed with the opportunity for practice, such that when the final assessment is taken they are all performing at or close to their asymptotic level. In the past, this has not been possible. Until recently, the only practical mechanism for giving people experience of testing has been through the use of practice tests or dissemination of information leaflets that provide examples of test items.

The Internet provides an ideal mechanism for making practice test materials available to people. Indeed, where testing is being used in this ‘formative assessment’ mode, the complex issues of test user authentication and supervision are of lesser importance, so long as the test content is different from that used for selection testing. For example, the SHL Direct for Students site\(^2\) provides sample items for verbal, numerical and diagrammatic tests as well as personality inventory items in both Likert-rating and ipsative (forced-choice) format. In addition, complete timed practice tests are available for potential test takers.

A range of other test publishers now provides Internet-based practice (though these generally do not include timed test items). Such systems provide the test taker with potentially useful feedback and are widely used to provide career advice, especially on web-based job boards.

Recruitment on the Internet

A major consequence of the rapid growth of the Internet and its accessibility is that increasing numbers of organisations are recruiting and selecting applicants for jobs online (see Bartram, 2000, and Lievens & Harris, 2003, for reviews of this area). Applicants for jobs and job-seekers are increasingly expecting to find work through the Internet rather than more traditional means. The Electronic Recruiting Index (ERI, 2000) showed a substantial increase in spending on e-recruiting in 1999. For 1998 the total was about $4.5 billion, while in 1999 it jumped to over $15 billion. The ERI forecasts steady growth from around $18 billion in 2000 to nearly $40 billion by 2005.

The Association of Graduate Recruiters in the UK (AGR, 2000) reported the results of a survey that showed that the number of recruiters recruiting online doubled in the year 1999 to 2000, from one-third to two-thirds of them. Nearly 90% of graduates were seeking their first jobs on the Internet and nearly 50% were applying online. What is more interesting is that employers report that the quality of applicants who applied online was higher than that of those who applied by traditional methods. Interestingly, the major change envisaged by the respondents as a consequence of growing use of the Internet was the

\(^2\) See http://www.shldirect.com/shldirect-homepage/SHLDirect-1.asp
demise of the hand-written application form. This reflects the growing trend to move away from the posting of CVs and resumes to the use of structured application forms systematically covering biographical data, experience, skills etc.

The Internet provides more than just a replacement for the traditional recruitment process. Online applicants can be responded to very quickly (within minutes by automated email, and within hours by recruiting managers). They can access sites where there is information about the company, the jobs it has, realistic job previews and so on. It is not just an online application process.

Internet recruitment has clear advantages for both applicants and recruiters. For recruiters, it provides the following:

- Larger applicant pools.
- Job profiling tools to set competency benchmark profiles against which to assess candidate fit.
- Very significant reductions in time to hire. Typically, times are reduced from four to eight weeks down to two weeks or less.
- Reduced cost per hire.
- The capability to email to candidates invitations to attend an interview or further assessment within hours of completing an application. This speed of response is likely to be critical for success in hiring for organisations competing for a limited talent pool.
- The potential for higher validity early sift leading to more cost-efficient selection.

For applicants, the advantages are:

- better feedback and advice on career and company choice
- good information about the available jobs and organisations
- access to a wider range of jobs and employers
- more rapid feedback and the ability to track one’s progress.

Well designed sites can help candidates make sensible decisions about whether they are qualified to apply and therefore increase the average suitability of the pool that go on to submit an application.

Candidate management and tracking are key features of Internet-based recruitment systems. These allow both candidates and hiring managers to know where a person is in the process and what their status is. People no longer have to post off an application form and then wait for weeks for a possible reply.

The key to making best use of the Internet for recruitment lies in using the new technology to apply valid objective assessment techniques to the initial sift process. Despite the fact the majority of large organisations now recruit over the Internet, most of them sift on the basis of purely demographic criteria and simple checks on relevant experience (Stanton & Rogelberg, 2001). What is
more, in a large number of cases they do this by printing out an applicant’s CV or resume, and then carrying out a traditional paper sift!

**Future Trends**

The future lies in developing structured assessments that can be completed online by job seekers and that can be shown to be job relevant. By doing this, it becomes possible to re-position online recruitment as a process of matching the competencies and capabilities of the applicant to the requirements of the job vacancy, and so produce a high quality shortlist.

Changes in technology need to be considered together with changes in recruitment practice. A major trend in personnel management has been the decentralisation of many operational responsibilities to staff at business unit, departmental or line-management level. A survey by the UK Institute for Personnel and Development (IPD, 1999) showed that in the UK line-managers are involved in determining recruitment criteria 97.4% of the time. The figures for central personnel staff (55.2%) and local personnel staff (36.6%) are much lower.

This trend to shift responsibility for recruitment and selection out to line management has implications for the design of selection systems. It is no longer safe to assume that a small number of highly trained personnel professionals will oversee the recruitment and selection procedures within their organisation. Thus, while needing to increase the sophistication of the recruitment and selection tools (to counter the increasing volumes of applicants available through online methods), we also need to ‘de-skill’ these tools from the point of view of the user. While it may be desirable, it is not practical to expect line managers in organisations to complete formal test user training courses before they begin to recruit personnel. Given such realities, the challenge for occupational psychologists is to design tools that are objective and job relevant but also easy and safe to use by relatively unskilled users.

**Post-Hire: Main Applications**

The most widely used application of the Internet for assessment within organisations is probably 360-degree feedback. There is now a wide range of systems available both for use over the Internet and for use on company Intranets. The major benefits are logistical. By its very nature 360-degree feedback is an administrative nightmare to manage. People involved in the process tend to be geographically dispersed but also need close supervision in order to ensure that the ratings are carried out to schedule and that sufficient raters are obtained for each focus of the assessment.

Good online systems focus on managing the workflow associated with the 360-degree process (Bartram et al., in press), from initial set-up and preparation of the people involved, through the management of the rating process.
(including delivery and scoring of questionnaires), to the production of reports and their delivery to feedback providers. Features of such systems will include:

- either candidate self-service or HR control over the administration
- users can choose their own raters
- task notifications, reminders and report delivery all automated
- either generic competency models can be used or they can be tailored to the organisation’s needs.

‘Good practice’ in carrying out a 360-degree feedback process can be built into the system by providing certain constraints on what users can do.

In addition to 360-degree feedback, there are a host of other applications where we are seeing the Internet becoming more widely used:

- organisational surveys
- multi-rater job analysis
- competency profiling
- individual development planning and tracking of development action plans
- performance appraisal
- performance management.

In addition, all the assessment procedures used in selection can be applied to post-hire development as well (e.g. through the use of virtual development centres).

FUTURE INTERNET APPLICATIONS

As noted earlier, there have been interesting developments in occupational testing within the military field and in the fields of education testing that have yet to impact significantly upon routine occupational testing. Certainly, we are starting to see developers and publishers move beyond traditional multiple-choice questionnaires (MCQs) and the use of classical test theory. There are a number of areas that are being developed.

Internet Interviews

Use is already made of the telephone for structured interviewing (Edenborough, 1994). However, the face-to-face interview serves a range of social functions other than the collection of information about the applicant (see, e.g., Herriot, 1989). One of the defining elements of a job interview is that it is an interactive dialogue between at least two people. It provides the opportunity for the applicant to learn about the potential employers and acts as a forum in which negotiation can take place between the parties.
Wide band G3 video-phone will provide a halfway house between the telephone interview and the ‘live’ face-to-face interview. Video-conferencing provides the employer with the opportunity to conduct single, pair or panel interviews without having the cost of transporting applicants to a common interview site. Certainly for overseas applicants, video-conference interviewing provides a major saving in cost and (for the applicant) time. By 2010 we can expect to see domestic digital TVs with built-in cameras being used as video-phones as part of their role as general-purpose multi-media entertainment and information centres. This will enable high fidelity interviewing to take place without applicants having to leave their homes.

It is likely that for certain jobs there will remain a final stage at which the job applicant and the employer need to meet face to face before entering into a formal employment contract. However, the role of this final meeting could shift away from that of an assessment process (as the information can be collected more efficiently online) towards that of discussing and agreeing the contract between applicant and employer. We could then see the interview become an event that occurs between the formal job offer being made and the applicant’s acceptance or rejection of it.

Reference Checks

It is already quite common to seek and transmit references by phone and by email. The use of the Internet to deliver structured and adaptive reference checks will add to the range of ways in which this information can be collected. It will also provide an effective means of providing a higher level of control over the administration of the reference-checking instrument. The same techniques can be used as are currently used in the systematic collection of information for 360-degree feedback.

Formal checking (subject to the necessary search and access permissions having been obtained) of medical, criminal and credit records will become very highly automated, as all the relevant data will be held on databases with (secure) Internet access.

Assessment and Development Centres

It is in the area of both group and individual assessment exercises that some particularly exciting new possibilities emerge. One of the earliest applications of the Internet (well before the advent of the World Wide Web) was for multi-user games such as Dungeons and Dragons. It is now possible to create multi-user exercises (e.g. business simulations) that can be closely monitored and assessed. The users need not be brought together to a single location, but could form part of a virtual assessment or development centre. While such procedures may have a greater part to play in training and development, they could also be used in a selection context.
For single-user exercises, web-based in-basket exercises are already available. These can be designed as relatively simple systems, for non-experienced users, or use software such as MS Outlook. In either case, people can be provided with emails, phone messages and background information, and have to work to obtain a set of objectives within some pre-defined constraints. The user can set tasks, make appointments, send emails and so on. All the actions and events can be logged, analysed and assessed.

The potential advantage of making such tasks Internet based is that it removes the geographical constraints on having to bring people together to take part in an assessment.

Cognitive Ability Testing

Excellent descriptions of the advances made in computer-based cognitive testing are provided in other chapters of this book and in reference books such as that by Wainer et al., (1990). For now we focus attention on the potential afforded by the Internet for remotely supervised cognitive ability testing. Use of the Internet for the deliver of cognitive ability tests is technically straightforward. Java or Flash applets, for example, can be written that provide high levels of control over the presentation of material and the timing of tests and responses. The use of downloaded applets also ensures that tests are not affected by denials of service from the user’s ISP occurring during a test session.

An example of an application of this new technology is provided by the SHL online Numerical Reasoning Test (Baron, Miles, & Bartram, 2001). The software developed for this test enables unique tests to be created for each individual, while ensuring that the difficulty level of the test is known and controlled. Standardised scoring is achieved through item calibration and then tests are constructed by constrained random generation from a large item bank. As a result of the fact that each test is unique, the issue of test security no longer arises. This test has been used by a number of organisations in an unsupervised mode as part of an initial screening sift in graduate recruitment. Research of the data subsequently collected from those candidates who proceeded to the second stage of the selection process showed high correlations with re-assessment data (collected under supervised conditions) and a substantial increase in the overall quality of the sifted applicants (twice as many passed the assessment centre as did before this sift tool was introduced).

PRACTICAL AND GOOD PRACTICE ISSUES

As the market for Internet-delivered computer-based testing develops, and as the technological sophistication of the products increases, so the issue of ensuring that those who use such tests and assessment tools follow good practice will increase in importance.
In this section we will consider a range of inter-related issues:

- Performance characteristics and technical limitations of the Internet as a test delivery medium: speed, network integrity, reliability, bandwidth etc.
- Security: protecting the publishers’ IPR; controlling test access and distribution, keeping scoring and rules confidential.
- Privacy: controlling access to test results, legal issues relating to data protection, privacy and storage.
- Fairness: equality of access for all groups to the net—closing the ‘digital divide’.

Performance

It is in the area of performance that the major current limitations of the Internet are to be found. Testing makes two main requirements of the delivery medium. First, it should provide the means of controlling the timing of delivery. Second, it should be robust and not fail mid-way through a test. Rapid delivery of pages to the user cannot be relied upon at present for a number of reasons. Many users are still on the end of a slow modem connection. While publishers can ensure that their servers deliver pages faster than users can call for them, slow connections at the user end, or other delays caused by the Intranet itself, can result in the actual appearance of pages on the user’s browser appearing to be very slow.

Even if the user has a broadband connection and there are no delays on the Internet, the most rapid delivery is only of value if it can be relied upon. Hang-ups and lost Internet connections can potentially terminate a test session in mid-stream. For some tests it is not practical simply to resume from the point at which a break occurred. For example, at present it is not safe to rely on the integrity of the Internet to the extent we would require for timed testing, even if we could be sure the page flow rate could be fast enough.

The easiest way to overcome these performance issues is to download any time critical material as an applet. This at least will ensure that the test administration is not dependent on the Internet for its timing and integrity. However, from the user’s point of view this may create another problem: if the connection is a slow one or the applet is large, the applet may take a considerable time to download. Furthermore, many organizations will not allow applets through their corporate fire-walls.

Consistency of Appearance

The Internet poses some of the same problems as stand-alone computer-based testing. For example, the test distributor has no direct control over the user’s screen size or resolution. For standalone systems software controls can be used to mitigate and control some of the extremes of variation in screen settings. For
browser-based testing, however, the level of control is rather less. Browsers are designed to leave the user in control of navigation around the Web and to be able to examine and modify the page display parameters in ways that we would wish to prevent in a normal test-taking situation.

Furthermore, there is no one ‘standard’ browser. Currently, two browsers dominate the market: Internet Explorer and Netscape Navigator. Unfortunately, these do not display information in exactly the same way. As a result, a test will look different and may behave differently depending upon the browser you are using. This problem is exacerbated by the fact that each of the browsers exist in a range of versions which also differ in how they render information.

Again, the solution to these problems, where they are likely to compromise the integrity of a test, is to create the test within an applet that can be downloaded and run on the user’s computer.

Security

Security concerns tend to be very high on the list of those worried about the use of the Internet for testing. The concerns over security need to be considered in relation to various sets of data:

- The test itself (item content, scoring rules, norms, report-generation algorithms, report content etc).
- The test taker’s identity—both authenticating the person’s identity and preserving confidentiality.
- Test results—ensuring that only those eligible to access the test scores are able to do so.

While all the above are areas of concern, it is important to put these into perspective by considering how the Internet, as a testing medium, compares with the current alternatives: paper-and-pencil and stand-alone computer-based testing.

Test Security

The key feature of the Internet is that, apart from the browser software itself, all the application software and all the data resides on the server, not on the user’s computer. Herein lie some of the main advantages of Internet based testing:

- All the important intellectual property associated with a test (scoring rules, norms, report-generation algorithms etc) remains on the server under the control of the distributor.
• This level of control provides the distributor with detailed knowledge about the use of their products: who is using what, and when. This has enormous potential commercial benefits.
• The test software and reference data set only exists in one location. This ensures that all users have access to the most up-to-date version. It also greatly simplifies the process of making changes, fixing ‘bugs’, updating norm tables and so on.

Authentication of Users

There is a range of levels of authentication that can be used. The distributor can either make a test open access, or exercise full control over who can access the test content, when they can access it and from where. Control can be exercised by requiring a username and password, or access can be limited to specific machines on the Internet (by only allowing those with particular IP addresses to use the test). By combining IP address checking with user passwords, a very high level of control can be exercised. Such a level of control was never possible for paper-and-pencil testing or for stand-alone computer-based testing.

While this level of authentication would be more than sufficient for managing access to personal banking information, it is not sufficient for ensuring that people are not cheating in a high-stake testing situation. For such a situation, a person could pass their identification to another person who would actually take the test, or they could have a group of helpers with them while they complete the test. Further advances in identification technology (fingerprint recognition and retinal eye-pattern recognition) would not really solve the problem of security in this sort of high-stake testing situation.

For this reason, it is likely that high-stake tests will continue to require the presence of a test administrator to confirm the identity of the test taker and ensure the test is completed under the correct conditions. While this is often noted as a disadvantage of Internet testing, it is really no different to other forms of testing technology.

Data Privacy

All the data generated by test takers resides on the central server. By applying best practice to the management of the server, the security of all aspects of the data can be far better assured than would be the case if the data were distributed amongst the various test users. While some people still have concerns about their data being held centrally, these concerns will decrease in the future as standards of security and good practice become more clearly established. In many ways worrying about having your data held on a professionally managed server is like worrying about your money being held in a bank—it is really a lot safer than keeping it at home!
Not only does the centralisation of data storage make it easier to manage but it also makes all test data potentially available for research and development purposes. This in turn can raise concerns in some people that they may be losing control over their data. Clearly, if the data are to be used for research and development purposes then this should be agreed with the data providers in advance. The service providers should have clear policies on how individual data are to be kept, for how long and who is allowed access to them. These details must be made clear to the test taker and be agreed to by the test taker before the data are collected. Of course, such policies must take account of, and be consistent with, national and international law on data privacy (for example, the European Union has a Directive relating to data privacy that is binding on all member countries of the Union).

**Fairness—and the Digital Divide**

There has been much concern expressed about the Internet creating a ‘digital divide’ between those with access to computer technology and those without (Keller, 1996). This is currently true on a geographical basis, with nearly all of the infrastructure and development of business taking place in North America, Europe and Asia-Pacific. This will change over the coming decade, but for some time we will not be able to use the Internet as the sole source of recruitment and selection in countries outside these three main areas.

However, if we consider just those areas where the infrastructure is well developed, does everyone have equal access to it? In considering any selection and recruitment process we need to consider its potential for adversely impacting on one or more particular groups within the population. From the viewpoint of litigation, the main ‘protected’ groups are ethnic minorities, women, and people with disabilities. More generally we should be concerned about equality of access in terms of geographical dispersion (rural versus urban), age, educational background and any other factors that may not be directly job relevant, but have an effect on access to the recruitment process.

In his review of Internet usage data, Bartram (2000) suggested that we are now seeing a second generation of Internet users and a move towards the equalisation of access. This view is supported by more recent data that shows a 35% growth rate from April 2000 to April 2001 in the number of African-American households online (Pastore, 2001a), bringing the total to over 50% of all such households. Pastore also quotes research showing that home Internet access for blue-collar workers surged by 52% in the year March 2000 to March 2001. In relation to age, the 55–64 age range experienced the second largest increase in the age demographic, growing by 20% to reach 52% of all households. Averages for the US in 2001 ranged from around 51 to 68% of all households for various demographic groups. Pastore (2001b) also quotes the results of a survey showing that Hispanics and Asians in the US are now more likely to be online at home that Caucasians and African-Americans, and that the growth rate for Hispanics has been over 80% in the past two years.
The first generation users are those who were there in the early days, who transitioned from pre-WWW Internet to the WWW in the mid-1990s and who are predominantly computer-skilled, young, white, male users. The second generation users are those who have come to use the WWW as it has become a part of the fabric of their work and home life (typically in the last two or three years). The latter generation are pretty equally divided between males and females, reasonably balanced in terms of ethnic mix, are older than first generation users and do not aspire to the higher levels of technical user-skill of the earlier generation. In terms of age, the main under-represented group is now unemployed people over 50.

TEST ADMINISTRATION OVER THE INTERNET

Exercising Control

Internet technology provides the opportunity for exercising much greater control over distribution of materials and intellectual property than traditional media. It provides us with the potential for:

- control over materials—immediate updating, ensuring that everyone is using the same versions
- control over prior practice—enabling test takers to start from a level playing field
- control over test takers—authentication: knowing who is taking the test
- control over test conditions—ensuring conformity to good practice.

However, it also provides us with the freedom to reduce or remove levels of control. Sometimes this may be desirable (e.g. providing large numbers of people with open access to good guidance on career choices), sometimes it may not (e.g. giving access to test materials that should be kept secure).

Much of the concern over Internet testing relates to issues of good practice. These concerns relate to three main areas:

- Ensuring that there is adequate control over the management of the assessment process.
- Ensuring that feedback and reporting is of high quality and contained within procedures that reflect good practice in assessment.
- Controlling the quality of tests delivered over the Internet.

Managing the process of assessment is a major topic and one that illustrates both how much control we can exercise and the dangers of not matching the levels of control to the requirements of the assessment process.
Management of the Assessment Process

Testing is a process involving a number of participants, each with differing roles. The exact nature and number of participants will vary depending on the nature of the test and the reason for testing. Typically the roles include:

- the initiator or ‘sponsor’ of the testing process
- the person responsible for managing the process
- the test administrator
- the test taker
- the person who will provide feedback to the test taker
- third parties who will be provided with information consequent upon the testing.

In addition to involving various people playing various roles, testing follows a sequence of events:

- the tests are chosen
- the arrangements are made for who is to be tested, when and where
- the tests are administered
- the scores are derived and reports generated
- the reports are delivered to the designated recipients
- feedback is provided to the test taker and/or relevant others.

The Internet provides the ideal medium for managing both the participants and the process. For example, some systems manage the process as a project that requires certain resources, in terms of people and materials, and has a time-line with a sequence of tasks and milestones. The workflow is managed using project templates that users can configure by entering the names of the various participants, selecting the instruments to be used and setting milestone dates for key points in the sequence of events that make up the testing process. The process is automatically managed by assigning tasks to people and communicating with participants by automated emails and hyperlinks. The level of control that can be exercised in this way over each participant is potentially configurable by the user. In this way, for example, it is possible to ensure that only qualified test users will have access to reports that require an understanding of a particular instrument or that only qualified test administrators are allowed to log test candidates onto the system.

Modes of Test Administration

Four modes of test administration have been defined (Bartram, 2001c). These modes form the basis for the guidelines on computer-based testing and the Internet developed by the ITC.
1. **Open mode.** These are conditions where there is no means of identifying the test taker and there is no human supervision. Examples of this include tests that can be accessed openly on the Internet without any requirement for test taker registration.

2. **Controlled mode.** This is similar to the open mode in that no human supervision of the test session is assumed. However, the test is only made available to known test takers. For the Internet this is controlled through the requirement for the test taker to be provided with a logon username and password.

3. **Supervised mode.** For this mode, a level of human supervision is assumed, whereby the identity of the test taker can be authenticated and test-taking conditions validated. This mode also provides a better level of control over dealing with unexpected problems or issues. For Internet testing, this mode is achieved by requiring the test administrator or proctor to log-in the candidate and to confirm that the testing was completed correctly at the end of the session.

4. **Managed mode.** This is a mode where a high level of human supervision is assumed and there is also control over the test-taking environment. For computer-based testing this is achieved through the use of dedicated testing centres. The organisation managing the testing process can define and assure the performance and specification of equipment in test centres. They can also generally exercise more control over the competence of the staff. In addition to standard ‘thin-client’ Internet applications, managed mode also provides the opportunity for delivering ‘thick-client’ applications under highly controlled conditions.

**Test Session Supervision Functions**

Supervision or proctoring has six functions.

1. Authenticating the identity of the test taker (i.e. establishing who is actually taking the test).

2. Establishing a positive rapport with the test taker (i.e. making sure that an appropriate climate is created for the test taking session and that the test taker is not unduly anxious).

3. Ensuring that instructions regarding standardised conditions are followed (e.g. making sure that timing conditions are adhered to, that calculators or other aids are used or not as instructed).

4. Dealing with unexpected conditions or problems that arise prior to or during the administration process (managing problems with equipment, hardware, disruptions during the test session, test taker disabilities etc).

5. Validating the test results (i.e. ensuring that the results obtained are what they appear to be, and were the product of the authenticated test taker operating unaided).

6. Ensuring that test materials are kept secure (i.e. making sure that no copies of test booklets or items are removed by the test takers).
The degree to which administration of an instrument requires the presence of a human supervisor will depend on the importance of direct supervision for each of these functions. This in turn depends on the nature and format of the test and the reasons why testing is taking place: for example, the type of test being administered (maximum versus typical performance); the format of the test (physical versus virtual); and the consequences of assessments (high versus low stakes).

One of the main reasons for requiring human supervision of testing is to manage the level of exposure that the item content has. Item generation techniques (Irvine & Kyllonen, 2002) provide us with the opportunity of developing a whole new range of tests for which this aspect of test security become less of a problem. This is a particular issue for high-stake tests where the item content needs to be re-used or where it might otherwise become known before the test session occurs. It is generally not an issue where tests of typical performance are concerned.

The management of test-taker honesty within a high-stakes assessment process is not just a matter of supervision. It is also a matter of the design of the whole process and the extent to which cheating or dishonest behaviour is likely to be detected. The assessment processes for job selection can be backed by an explicit ‘honesty policy’ to which candidates are asked to sign up. This is supported by the process of re-assessing, under controlled supervised conditions in the final stages of the selection process, any key competencies which formed the basis for the sift. Such agreements are used to ‘click-wrap’ the application. While such contracts are not legally binding or able to guarantee that the applicant has actually abided by them, they do help provide a clear set of expectations and explain that failure to abide by these conditions could have undesirable consequences.

At present we have very little hard data to show the impact of such approaches on test taking strategies. What we do know, however, is that the use of an unsupervised randomly generated numerical reasoning test during the sift stage of a recruitment procedure can dramatically improve the quality of the candidates who pass that sift (Baron et al., 2001).

Feedback and Reporting

Just as it is necessary in some conditions to ensure that there is a human test administrator or proctor present to ensure that high-stake assessments are carried out properly, so there will also be conditions where it is important to ensure that feedback is provided to a test taker by a qualified person rather than over the Internet. The question of when this is necessary is a matter of professional judgement. Generally, one can argue that in any situation where the feedback is complex and needs careful explanation face-to-face feedback should be given. An in-between option is to provide simple feedback online with a phone-in ‘help-line’ for people to get more in-depth feedback.
Most computer-generated test reports are designed for the test user rather than the test taker (Bartram, 1995a). Considerable care and attention needs to be given to reports that are intended to provide the sole source of feedback for the test taker.

In practice, the situations where feedback needs to be provided on a face-to-face basis will tend to be the same ones as where the assessment itself needs to be supervised. As such, providing for this is no more of a problem than it would be for traditional paper-and-pencil testing. With well designed Internet testing process-management software, the logistics of arranging for test sessions and feedback appointments are much simpler than for traditional assessment.

Test Quality

The effect on a test’s psychometric properties of delivering it over the Internet must be considered. Examples of bad practice abound. For example, some people have taken timed, supervised, paper-and-pencil tests and put them onto the Internet as un-timed and unsupervised. Clearly, in such cases, one cannot regard the Internet version as the ‘same’ test as the original.

In general, when a test is presented in some medium other than the one in which it was developed, it is necessary to check the equivalence of the new form. In practice this is most likely to be an issue for timed ability and aptitude tests. Most research suggests that the data obtained from un-timed self-report inventories are not affected by whether the test is administered on paper or on computer (see Bartram, 1994; Bartram & Brown, 2004; Mead & Drasgow, 1993; Salgado & Moscoso, 2003).

In summary, the main concern over the use of unsupervised modes of test administration is that such administration will adversely affect validity and, therefore, utility. In addition, lack of supervision can result in the compromising of test security. This may be a critical issue for traditional ability tests, which have a fixed set of items. Once these become widely known, the test will be of little value. However, this is far less of an issue for self-report personality measures or for tests where item content is continually modified.

CONCLUSIONS

It has been argued that the Internet actually allows us to exercise far more control than we have been able to do in the past over distribution of materials, management of the assessment process and the collection of data.

The key advantage, as the medium matures, will lie in test producers and publishers being able to assume the availability and accessibility of a ubiquitous infrastructure through which to deliver new products and services. Test users and test takers will have access to a wider range of services, better matched to their needs and better supported. Test designers will be able to
consider new possibilities for assessment design: real-time interactive virtual group exercises using emails or videophone conferencing; realistic in-tray tasks and so on.

The advantages of the Internet are also its dangers. Anyone can now set up a home-page and ‘publish’ a test. Assessment authoring systems are already available for producing and delivering simple tests and questionnaires on the Web. Dozens of ‘tests’ can be found that provide interesting looking reports (a quick search of the web for measures of Emotional Intelligence, for example, found more than a dozen questionnaires). However, there is typically no indication of the quality of these and the unwary user can be forgiven for failing to distinguish between serious assessment and trivia.

From the test user and test taker’s points of view, it is becoming increasingly difficult to discriminate between good tests and bad. In testing, the medium is not the message, as the quality of the test is always hidden in the technical data. As a result, the emphasis placed by the major publishers on technical and ethical standards and good practice will become increasingly important.

We need to consider the implications of the new technology on standards and good practice in assessment (Bartram, 1995b, 1998, 2001a). In particular, though, we need to do this at an international level. National professional associations and national publishers can no longer operate as closed systems. The presence of international networks, globalisation of industry and communications means that testing is now an international activity and individual nations need to be prepared to work as open systems within agreed international standards frameworks.

In the next few decades we will see the availability of Internet technology with which one can create fully immersive virtual realities (sound, touch, vision and smell) for single- and multi-person assessments. The opportunities this will provide for assessment are almost without limit. However, with each new step advance in technology come associated new issues for best practice.

The work of the International Test Commission on test adaptation (Hambleton, 1994; Hambleton, Merenda & Spielberger, 2005; International Test Commission, 2001a) and test user guidelines (Bartram, 1998, 2001b; International Test Commission, 2001b) provides a valuable starting point for future developments at the international level. However, more needs to be done if standards are to keep pace with the changes in technology, and those involved in recruitment and selection procedures are to be protected from bad practice and poor assessment tools. The International Test Commission embarked on a new project in 2001 to develop guidelines for computer-based testing and testing on the Internet. The ITC Conference in Winchester in 2002 and the resulting chapters in this book have been major contributions to this development. In the time it has taken to develop this book, these guidelines have also been developed. They have been through a number of cycles of international consultation and will be published, as befits guidelines on this subject, on the International Test Commission website: http://www.intestcom.org by the time this chapter is published.
We are in the middle of a very exciting revolution. We need radically to review our conception of assessment as a process and to reconsider the relationships between the various stakeholders in the assessment process: test developers, test publishers, test users, test takers, consumers of test results, professional bodies and lawmakers. While the essential principles of best practice will not change, as they are independent of assessment technologies, actual standards do need to be reviewed and re-considered in terms of the relationships between virtual tests and roles in cyber-space, rather than material tests and people in real geographical space.

REFERENCES

Association of Graduate Recruiters (AGR). (2000). Going to work on the Web: Web-based graduate recruitment (AGR Briefing No. 11). Warwick, UK: AGR.


