Computer-Based Testing and the Internet
Computer-Based Testing and the Internet
Issues and Advances

Edited by

Dave Bartram
SHL Group plc, Thames Ditton, Surrey, UK

Ronald K. Hambleton
University of Massachusetts at Amherst, USA

John Wiley & Sons, Ltd
Contents

About the Editors vii
List of Contributors ix

Introduction: The International Test Commission and its Role in Advancing Measurement Practices and International Guidelines 1
Thomas Oakland

1 Testing on the Internet: Issues, Challenges and Opportunities in the Field of Occupational Assessment 13
Dave Bartram

2 Model-Based Innovations in Computer-Based Testing 39
Wim J. van der Linden

3 New Tests and New Items: Opportunities and Issues 59
Fritz Drasgow and Krista Mattern

4 Psychometric Models, Test Designs and Item Types for the Next Generation of Educational and Psychological Tests 77
Ronald K. Hambleton

5 Operational Issues in Computer-Based Testing 91
Richard M. Luecht

6 Internet Testing: The Examinee Perspective 115
Michael M. Harris

7 The Impact of Technology on Test Manufacture, Delivery and Use and on the Test Taker 135
Dave Bartram

8 Optimizing Quality in the Use of Web-Based and Computer-Based Testing for Personnel Selection 149
Lutz F. Hornke and Martin Kersting
CONTENTS

9  Computer-Based Testing for Professional Licensing and Certification of Health Professionals  163
    Donald E. Melnick and Brian E. Clauser

10  Issues that Simulations Face as Assessment Tools  187
    Charles Johnson

11  Inexorable and Inevitable: The Continuing Story of Technology and Assessment  201
    Randy Elliot Bennett

12  Facing the Opportunities of the Future  219
    Krista J. Breithaupt, Craig N. Mills and Gerald J. Melican

Index  253
About the Editors

Dave Bartram is Past President of the International Test Commission and is heading ITC projects on international guidelines for standards in test use and standards for computer-based testing and the Internet. He is Chair of the British Psychological Society’s Steering Committee on Test Standards and Convenor of the European Federation of Psychologists’ Associations Standing Committee on Tests and Testing. He is President-Elect of the IAAP’s Division 2.

Professor Bartram is Research Director for SHL Group plc. Prior to his appointment with SHL in 1998, he was Dean of the Faculty of Science and the Environment, and Professor of Psychology in the Department of Psychology at the University of Hull. He is a Chartered Occupational Psychologist, a Fellow of the British Psychological Society (BPS) and a Fellow of the Ergonomics Society. In 2004 he received the BPS award for Distinguished Contributions to Professional Psychology. His specialist area is computer-based testing and Internet assessment systems. Within SHL he is leading the development of their next generation of Internet-based delivery systems and the development of a multi-dimensional generic Competency Framework.

He has published large numbers of popular, professional and academic articles and book chapters, and has been the Senior Editor of the BPS Test Reviews. He has been an editor or co-author of several works including the 1992, 1995 and 1997 BPS Reviews of Psychometric Tests; Organisational Effectiveness: the Role of Psychology (with Ivan Robertson and Militza Callinan, published in 2002 by Wiley) and the BPS Open Learning Programme for Level A (Occupational) Test Use (with Pat Lindley, published by BPS Blackwell in 1994).

Ronald K. Hambleton holds the title of Distinguished University Professor and is Chairperson of the Research and Evaluation Methods Program and Executive Director of the Center for Educational Assessment at the University of Massachusetts, Amherst, in the United States. He earned a B.A. in 1966 from the University of Waterloo in Canada with majors in mathematics and psychology, and an M.A. in 1967 and Ph.D. in 1969 from the University of Toronto with specialties in psychometric methods and statistics. Professor Hambleton teaches graduate-level courses in educational and psychological testing, item response theory and applications, and classical test theory.
models and methods, and offers seminar courses on applied measurement topics. He is co-author of several textbooks including (with H. Swaminathan and H. Jane Rogers) Fundamentals of Item Response Theory (published by Sage in 1991) and Item Response Theory: Principles and Applications (published by Kluwer in 1985), and co-editor of several books including International Perspectives on Academic Assessment (with Thomas Oakland, published by Kluwer in 1995), Handbook of Modern Item Response Theory (with Wim van der Linden, published by Springer in 1997) and Adaptation of Educational and Psychological Tests for Cross-Cultural Assessment (with Peter Merenda and Charles Spielberger, published by Erlbaum in 2005). His research interests are in the areas of item response model applications to educational achievement and credentialing exams, standard-setting, test adaptation methodology, score reporting and computer-based testing. He has received several honors and awards for his more than 35 years of measurement research including honorary doctorates from Umea University in Sweden and the University of Oviedo in Spain, the 1994 National Council on Measurement in Education Career Award, the 2003 Association of Test Publisher National Award for Contributions to Computer-Based Testing, and the 2005 E. F. Lindquist Award for Contributions to Assessment. Professor Hambleton is a frequent consultant to state departments of education, national government agencies and credentialing organizations.
Contributors

Dave Bartram, SHL Group, The Pavilion, 1 Atwell Place, Thames Ditton, Surrey, KT7 0NE, UK

Randy Bennett, Educational Testing Service, Rosedale Road, MS 17-R, Princeton, NJ 08451, USA

Krista J. Breithaupt, American Institute of Certified Public Accountants, Harborside Financial Center, 201 Plaza Three, Jersey City, NJ 07311-3881, USA

Brian E. Clauser, National Board of Medical Examiners, 3750 Market Street, Philadelphia, PA 19104-3190, USA

Fritz Drasgow, Department of Psychology, University of Illinois, 603 East Daniel Street, Champaign, IL 61820, USA

Ronald K. Hambleton, University of Massachusetts at Amherst, Center for Educational Assessment, Hills South, Room 152, Amherst, MA 01003, USA

Michael Harris, College of Business Administration, University of Missouri, 8001 Natural Bridge Road, St Louis, MO 63121-4499, USA

Lutz Hornke, University of Aachen, Department of Psychology, Aachen Technical University, D-52056 Aachen, Germany

Charles Johnson, Competence Assurance Solutions Ltd, 9 Lawford Road, Rugby, CV21 2DZ, UK

Martin Kersting, University of Aachen, Department of Psychology, Aachen Technical University, D-52056 Aachen, Germany

Richard Luecht, Curry Building 209/ERM, University of North Carolina at Greensboro, P.O. Box 26170, Greensboro, NC 27402-6170, USA

Krista Mattern, Department of Psychology, University of Illinois, 603 East Daniel Street, Champaign, IL 61820, USA
Gerald J. Melican, American Institute of Certified Public Accountants, Harborside Financial Center, 201 Plaza Three, Jersey City, NJ 07311-3881, USA

Donald Melnick, National Board of Medical Examiners, 3750 Market Street, Philadelphia, PA 19104-3190, USA

Craig Mills, American Institute of Certified Public Accountants, Harborside Financial Center, 201 Plaza Three, Jersey City, NJ 07311-3881, USA

Tom Oakland, University of Florida at Gainesville, 1403 Norman Hall, Gainesville, FL 32611-7047, USA

Wim J. van der Linden, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
We reside in various communities and live in one world. Although we may reside in Beijing, China, on a kibbutz in Israel, or in Muleshoe, Texas, we are aware of the impact of world events on our lives.

This has not always been true. Throughout most of history, life generally was impacted by dominant qualities in one’s community or a region reachable within one day. Important events that impacted the lives of those living thousands of miles away generally were unknown and had little impact outside that immediate area. Now, however, knowledge of events that occur throughout the world can be received in real time and can impact others’ lives greatly.

Electronic technology accounts for much of this change. The use of telephones, radios, televisions, and computers has drawn people who live hundreds, even thousands of miles away into a common neighborhood by enabling them to have improved access to information, establish and maintain relationships, and in other ways engage in activities beyond their communities and nations.

---


*Computer-Based Testing and the Internet: Issues and Advances.*
People increasingly are never away from their phones. Cameras transmit pictures showing the flow of cars and people. Credit card use requires digitized information be sent to satellites and then returned to the earth, perhaps to a location thousands of miles from where the card was used. TVs convey pictures of warfare in real time. The use of technology is pervasive.

Computer use may account for more change during the last two decades than any other form of technology. Our use of computers has shaped the ways we work, attend school, bank, invest, pay taxes, acquire and disseminate information, plan vacations, order food, and date—to name only some of the more obvious areas.

The development, use, and availability of any technology, to be acceptable, must serve the public good and should assist professionals, as needed, in their efforts to serve the public.

Change that follows the introduction of new technology may be beneficial or harmful to society. Technology generally is intended to enhance the quality of life for a large number of people by providing needed services at lower costs. However, if unregulated and used inappropriately, technology can adversely impact lives. For example, the viewing of violence and sex by children and youths contributes to behaviors and attitudes that often are narcissistic and do not serve the public. Unregulated technology can be harmful. Thus, efforts are needed to help insure technology serves people well.

Technology also should serve professionals well. The introduction of new technology impacts the manner in which professionals can be expected to conduct their work. Professionals incorporate various forms of electronic technology into their practices when they increase effectiveness and efficiency. Their use of technology is preceded by study that informs them of the best ways to shape and apply this technology.

Tests constitute some of psychology’s most important technology. Their use is universal, often starting with newborns and extending through the elderly. Their use is intended to serve the public by improving the ability to describe important behaviors, identify personal strengths, diagnose disabling conditions, assist in making administrative decisions, and help predict future behaviors. However, test use, if unregulated and used inappropriately, can adversely impact lives.

Computer use is shaping the ways tests are constructed, normed, validated, administered, and scored. The reach of this technology on test practices is broad and international. Thus, involvement of organizations that transcend one community, even one nation, is needed to envision, create, promote, regulate, and in other ways assist in forming and reforming test-related services that serve the professions responsible for test development and use as well as the public.

Many issues pertaining to test development and use are international in scope and thus need to be addressed at this level. The International Test Commission, International Association for the Evaluation of Educational Achievement, International Association of Applied Psychology, International Union of Psychological Sciences, Organization for Economic Cooperation and
Development, World Bank, and World Health Organization are among those providing leadership internationally. The work of the International Test Commission is becoming particularly prominent in this endeavor and is summarized below.

This introductory chapter reviews some historical features of test development and use and highlights the role of the International Test Commission in furthering these efforts. Thus, a goal of this chapter is to describe the context for forming the International Test Commission, initiating the 2002 Winchester ITC Conference on Computer-Based Testing and the Internet, as well as the contents of this book.

**SOWING THE SEEDS OF TEST DEVELOPMENT AND USE**

The first widespread use of tests occurred in China more than 3000 years ago. Measures of problem solving, visual spatial perception, divergent thinking, and creativity were used somewhat commonly. Later, under the Sui dynasty (581–618), a national civil service examination system was established that assessed three broad and important areas: cultural knowledge, planning and administration, and martial arts (Wang, 1993). Forms of this examination system continued in China to the end of the 19th century. However, few if any other countries seemingly duplicated these assessment practices or developed others on a national scale until the 20th century. A discipline of psychology, devoted to the study of individual differences in human behavior, was needed to initiate and sustain advocacy for these developments.

Psychological science first emerged from laboratories established by Fechner, Weber, and Wundt in Germany, by Galton in England, and by other pioneers who helped establish the scientific foundation for this fledging discipline. These early efforts to develop and use various tests and other measures, largely for research purposes, set the stage for later efforts that lead to the creation of psychometrics and other test-related specializations within psychology.

Before World War II, leadership in psychology rested among those who saw psychology as an academic discipline, one that was not sufficiently prepared to offer professional services at the same level as well established professions. Tests were developed and used mainly to conduct research and assess educational attainment. Thus, most psychology departments initially resisted pressure to offer programs that prepared practitioners to develop and use tests.

World War I provided one of the first large scale opportunities to evaluate the use of tests in applied settings. Within the United States, group tests of mental abilities were developed, found to be useful in selecting soldiers, and found to be psychometrically sound. Later efforts to develop other group and individually administered tests relevant to issues in psychology and education also proved to be effective. World War II expended the scope of test use, thus supporting its use in the selection, training, and placement of military recruits. By the 1950s, the viability of applied uses of tests was becoming widely accepted.
Following WWII, testing technology was new and generally unregulated. As noted above, the development, use, and availability of any technology, to be acceptable, must serve the public good and should assist professionals, as needed, in their efforts to serve the public. However, during the late 1940s and early 1950s, standards for test development and use had not been developed, and ethical issues regarding test use had not been addressed. Thus, there was little assurance that the emerging testing technology would serve the public and assist professionals.

THE EMERGENCE OF TEST DEVELOPMENT, PURCHASE, AND USE AND THUS THE NEED FOR REGULATIONS

During the 1960s, Swiss psychologists resembled their colleagues in many nations. Most worked in universities and had little interest in or commitment to preparing students for applied careers in psychology. Some believed the discipline of psychology had not matured sufficiently to warrant the professional practice of psychology and its use of tests. Thus, lacking opportunities to acquire needed knowledge and skills within universities, applied psychologists interested in test use often were self-taught or took courses elsewhere. Although separate Swiss psychological associations were formed for those who had scholarly and applied interests, they united to address two common and inter-related issues: a desire to protect the title psychologist as well as methods psychologists developed and used, including tests. At that time, anyone could purchase tests. The Swiss expressed considerable concern when learning that the Rorschach Inkblot Test, one developed by a Swiss, as well as other assessment measures, could be purchased without restrictions in bookstores by persons unqualified to use them.

In 1968, Jean Cardinet, a Swiss psychologist, discussed applications of ethical standards in testing at the Swiss Psychological Society. He and others were concerned with the use of important tools to make life-changing decisions by people who lacked sufficient training and experience. In addition, tests were used that lacked adequate psychometric qualities and diagnostic value. Cardinet together with his Swiss colleagues discussed possible solutions to these test-related difficulties. Although these problems were apparent in Switzerland, they were also apparent in other countries. For example, a restriction on the sale of tests in Switzerland would be ineffective if the same tests could be purchased in neighboring countries. Thus, regional and international remedies were needed to address this and other test-related issues.

Cardinet and his colleagues saw the need to create test commissions in all countries charged with two responsibilities: to decide who is authorized to use various types of test and to scrutinize their sales to insure they do not bypass needed controls (Cardinet, 1974a, 1974b, 1974c, 1975, 1995). Cardinet envisioned each national commission designing a program of work in light of its national needs and conditions and cooperating with other national commissions through their association in an international test commission.
Upon contacting all national psychological societies, Cardinet became aware that problems experienced by his Swiss colleagues were international in scope and that some countries addressed them better than others. Fifteen countries had established test commissions, and additional countries expressed interest in forming one. Support for an international association devoted to test-related issues was also expressed.

**FORMATION OF THE INTERNATIONAL TEST COMMISSION**

Cardinet expressed hope that an international test commission would develop an ethics code and standards for test construction and evaluation, create a journal to promote an exchange of information, and assist colleagues working in developing countries to improve conditions governing the development and use of tests.

In 1974, leadership for forming a fledging international test commission transferred from Cardinet to Ype Poortinga, a leader of the Dutch Test Commission. During the next four years, Poortinga’s efforts were directed toward developing an administrative infrastructure, including drafting a constitution, establishing a newsletter, and forming an advisory committee. A constitution for the International Test Commission (ITC) was approved in 1976, at which time Poortinga assumed the office of president. Articles of Incorporation for the International Test Commission, Inc. were filed in 1999 as a not-for-profit corporation within the state of Florida, thus establishing the ITC as a legal entity.

**International Test Commission’s Primary Goals**

The ITC’s primary goal has been to assist in the exchange of information on test development and use among its members and affiliate organizations as well as with non-member societies, organizations, and individuals who desire to improve test-related practices. Consistent with Cardinet’s vision (i.e. to create an ethics code and standards for test construction and evaluation and a journal to promote an exchange of information, and assist colleagues in developing countries to improve conditions governing the development and use of tests), the ITC has worked to promote its goals through expanding membership, working cooperatively with other organizations, developing and promoting guidelines for test development and use, and engaging in other forms of publications and communication.

**International Test Commission Membership**

The ITC’s constitution initially identified two membership categories: full members consisting of national test commissions recognized by a national
psychological association and affiliate members consisting either of international associations that had an interest in assessment or national groups from countries not full members of the ITC. Revisions to the constitution expanded membership to include testing companies, universities, organizations interested in tests including those unable to pay yearly dues (e.g. psychological associations from developing countries), and individuals involved in test development and use. The increased breadth of membership reflects the ITC’s desire to further increase and broaden its membership base and to include as members all persons with legitimate interests in test development and use, including non-psychologists.

As of March 2003, the ITC had 127 members. Most national psychological societies in Europe and North America, a number of national societies from other continents, nearly all major test publishers, together with many research departments in educational and psychological measurement are ITC members.

Working Cooperatively with Other Organizations to Achieve Common Goals


Development and Promotion of Guidelines for Test Development and Use

ITC’s leadership in developing and promoting guidelines for test development and use became more evident in the early 1990s, when it began sponsoring international and regional meetings on topics that hold special
importance to test issues that had a decided international scope. Five ITC-sponsored conferences are described below.

1993 Conference on Test Use with Children and Youth: International Pathways to Progress

The first international conference independently proposed and organized by the ITC was held in June 1993 at St. Hugh’s College, Oxford University, in England (Oakland & Hambleton, 1995). The conference’s primary focus was on testing practices of children and youth in developing countries. This focus was consistent with the goals of the World Summit for Children in 1990 and the World Conference on Education for All in 1991. Both underscored the importance of promoting children’s educational development though using test results to assist in planning and evaluation efforts. The ITC conference underscored its concerns about a need to improve educational and psychological assessment practices for children and youth. This need was especially apparent in developing countries as underscored in a series of publications (Hu & Oakland, 1991; Oakland & Hu, 1991, 1992) on the status of test development and use in 44 countries.

Guidelines for Adapting Tests

Test use in many countries is characterized by translating tests originally developed elsewhere. This common practice is contrary to common sense and violates laws, ethics codes, and psychometric principles.

The ITC observed a growing interest in avoiding problems associated with the use of translated tests and in pursing proper adaptation of educational and psychological tests. For example, by 1992, some tests developed in the United States had been translated and adapted into more than 50 languages. At the same time, efforts to adapt tests were hindered, in part, because technical guidelines governing test adaptations were unavailable. The ITC, under the leadership of Ronald Hambleton and with the assistance of an international panel, developed guidelines for adapting educational and psychological tests (e.g. Hambleton, 1994; Muniz & Hambleton, 1997; van de Vijver & Hambleton, 1996; Hambleton, Spielberger, & Merenda, 2005; also see the ITC web site).

1999 Conference on Test Adaptations: Adapting Tests for Use in Multiple Languages and Cultures

The availability of the ITC guidelines on test adaptations, advances in test translation methodology, and considerable international interest in the topic warranted the International Conference on Adapting Tests for Use in Multiple Languages and Cultures, held in May 1999 at Georgetown University in Washington, DC (Hambleton, Spielberger, & Merenda, 2005).
1999 Conference on Cultural Diversity and European Integration

A conference on Cultural Diversity and European Integration, held at the University of Graz, Austria, in June 1999, also featured issues associated with test adaptations. The ITC co-sponsored this pan-European event with the International Association for Cross-Cultural Psychology. Primary motivation for the conference was the widespread immigration taking place throughout Europe and the need for better understanding cross-cultural issues as the European integration process continued.

2000 International Congress of Psychology in Stockholm

The International Union of Psychological Sciences convenes the International Congress of Applied Psychology every four years. It also sponsors advanced research training seminars in conjunction with the congresses. The ITC presented a seminar on psychological test adaptations to diverse cultures and measuring personality cross-culturally during its 2000 meeting in Stockholm (Oakland & Lonner, 2001).

2002 Conference on Computer-Based Testing and the Internet

The ITC leadership recognizes the value of drawing together people with common interests to discuss topics and issues important to test development and use. Somewhat small and well focused conferences can allow everyone to participate, promote an exchange of information, promote lasting relationships, and result in other positive unintended outcomes.

The ITC-sponsored conference on Computer-Based Testing and the Internet was designed with these beliefs. Directed by Dave Bartram and held in Winchester, England, in June 2002, the 250 conferees discussed issues relating to the uses and misuses of electronically transmitted tests and other assessment devices together with methodological, technical, legal, ethical, and professional issues arising from applications of computer-based technology to testing. This conference together with the contents of this book and new Guidelines for Computer-Based Testing and Testing on the Internet (Bartram & Coyne, 2003) launches the ITC’s efforts to promote knowledge and understanding of uses and possible abuses in using computers to assist in test development and use.

International Guidelines for Test Use

The ITC’s commitment to promote practices that can benefit test use is seen in its original charge. Early records reveal uneasiness as to the presence of unqualified persons using tests, their making important decisions despite having limited preparation and experience, and their use of tests that lack suitable norms and sufficient validity.
The ITC, under Dave Bartram’s leadership, developed guidelines for the fair and ethical use of tests (Bartram, 1998, 2001). These guidelines provided an internationally agreed framework from which standards for professional preparation and test user competence and qualifications can be derived. The guidelines were approved by the ITC in 1999 and have been endorsed by the European Federation of Psychologists Associations Standing Committee on Tests and Testing. Psychological societies in a number of countries, including Argentina, Brazil, Croatia, Denmark, France, Germany, Lithuania, the Netherlands, Norway, Portugal, Slovenia, and Sweden, have translated the guidelines for use. The guidelines for the fair and ethical use of tests contribute importantly to the previously approved guidelines for test adaptations.

Publications and Communications

A Newsletter

The need to establish a journal to help promote communication and dissemination internationally has been a long-standing goal. Progress toward this goal was slow but continuous. The Newsletter served as the ITC’s first vehicle of communication and dissemination. The Newsletter later spawned the ITC Bulletin, a publication that included articles of a scholarly nature as well as more general membership information. In 1991, the ITC decided to separate the two content areas by developing a more scholarly outlet and re-establishing a newsletter.

Association with the European Journal of Psychological Assessment

The ITC immediately entered into an informal agreement with the European Journal of Psychological Assessment that resulted in reserving 16–20 pages of each issue as an ITC-sponsored publication. This relationship continued through the 1999 volume years.

Testing International

The newsletter was re-established, renamed Testing International, and made available through the ITC website. Anita Hubley served as its editor through to 2005, when she was succeeded by Jan Bogg.

International Journal of Testing

A long-standing goal to create an ITC-sponsored journal was achieved in 1999. The International Journal of Testing was launched, with Lawrence Erlbaum Associates, Inc., Publishers. Norbert Tanzer and then Bruno Zumbo served as its
editor through to 2005, when he was succeeded by John Hattie. The *International Journal of Testing* enables the ITC to meet a long-standing need: to create a truly international journal that addresses issues important to test development and use. The creation of the *Journal* is consistent with other efforts to positively impact testing internationally.

**ITC Website**

The creation of a website ([http://www.intestcom.org](http://www.intestcom.org)) provided another method to improve communication between members and to communicate ITC’s goals and programs internationally. The site provides information about membership, ITC projects, and *Testing International*, and launches documents developed and promulgated by the ITC.

**ITC Contributions Through Books**

The ITC’s long-standing commitment to disseminate information also is found in its members’ publishing books as well as sponsoring and presenting papers at national, regional, and international meetings. The first book highlighted advances in educational and psychological testing (Hambleton & Zaal, 1991). The second book focused on issues underscored in the Oxford conference, namely test use with children and youth (Oakland & Hambleton, 1995). Issues central to test adaptations were discussed by Hambleton, Spielberger, and Merenda (2005). Special issues of the *ITC Bulletin* that appeared in the *European Journal of Psychological Assessment* featured topics on computers in psychology in 1994, advances in assessment practices in 1997, and the ITC guidelines for adapting educational and psychological tests in 1999.

*Computer-Based Testing and the Internet: Issues and Advances*

This current book, edited by Bartram and Hambleton, represents another step forward, given the ITC’s mission to advance test development and use internationally. Its contents reflect the leadership of the co-editors together with the continued dedication by the ITC to engage in scholarly activities intended to help promote test development and use internationally. It builds upon and extends contributions derived from the ITC-sponsored 2002 Winchester conference on Computer-Based Testing and the Internet.

A key theme of the Winchester conference was to provide the background to the development of guidelines on computer-based testing and testing on the Internet. The development of these guidelines has progressed in parallel with the production of this book and we expect both to be published at about the same time. These guidelines, like all the other ITC guidelines, can be found on the ITC’s website.
REFERENCES


CHAPTER 1

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

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 hyperlinked 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