



# TESTING INTERNATIONAL

*Editor: Anita M. Hubley*

**PRESIDENT**

Professor David Bartram  
SHL Group PLC  
The Pavilion  
1, Atwell Place  
Thames Ditton  
Surrey, KT7 0SR  
United Kingdom

**PRESIDENT-ELECT**

Professor Bruce Bracken  
College of William and Mary  
School of Education  
P.O. Box 8795  
Williamsburg, VA 23187-8795  
U.S.A.

**SECRETARY**

Professor Jacques Grégoire  
Université catholique de Louvain  
Faculté de Psychologie  
Place du Cardinal Mercier, 10  
1348-Louvain-la-Neuve  
Belgium

**TREASURER**

Professor Em. Barbara M. Byrne  
School of Psychology  
University of Ottawa  
145 Jean-Jacques Lussier  
Ottawa, Ontario  
Canada K1N 6N5

**PAST-PRESIDENT**

Professor Thomas Oakland  
College of Education  
University of Florida  
P.O. Box 117047  
Gainesville, FL 32611-7047  
U.S.A.

**COUNCIL MEMBERS**

Professor J. Muñiz  
Oviedo University, Spain  
Professor Cheryl Foxcroft  
University of Port Elizabeth,  
South Africa  
Ms. Patricia Genat  
NCS, Australia  
Dr. Marise Born  
Erasmus University Rotterdam,  
The Netherlands  
Dr Iain Coyne  
University of Hull, UK

**REPRESENTATIVES**

Professor Ronald K. Hambleton  
(IAAP)  
University of Massachusetts, U.S.A.  
Professor John G. Adair (IUPsyS)  
University of Manitoba, Canada

**EDITORS**

*International Journal of Testing*  
Professor Bruno D. Zumbo  
Dept. of ECPS  
2125 Main Mall  
University of British Columbia  
Vancouver, B.C.  
Canada V6T 1Z4

**Testing International**

Dr. Anita M. Hubley  
Dept. of ECPS  
2125 Main Mall  
University of British Columbia  
Vancouver, B.C.  
Canada V6T 1Z4

## EDITOR'S NOTE

2

## PRESIDENT'S LETTER

2

## BRIEF ARTICLES

- Testing on the Internet: Issues, Challenges and Opportunities  
*Dave Bartram (U.K.)* 4
- Ten Ethical Issues Associated with Computer-based Testing and the Internet  
*Thomas Oakland (U.S.A.)* 6
- Contemporary Uses of Computer- and Internet-based Testing, a Particularly Appealing Context for the Application of Bayesian Statistical Methods  
*Bruno D. Zumbo (Canada)* 8
- BS 7988: A New British Standard on using IT in Delivering Assessments  
*John Kleeman (U.K.)* 10

## THE NOTICEBOARD

- Barbara Byrne Receives Jacob Cohen Distinguished Teaching/Mentoring Award 11

## UPCOMING CONFERENCES ON TESTING AND ASSESSMENT

11

**OFFICIAL NEWSLETTER**

**OF THE INTERNATIONAL TEST COMMISSION**

## EDITOR'S NOTE

Anita M. Hubley  
 University of British Columbia  
**CANADA**

Dear Friends,

One of the suggested topics in the graduate seminar on *Test Development* that I taught in the measurement program at the University of British Columbia this past winter was 'Internet testing'. Internet testing is not my particular area of expertise so both my students and I were eager to discuss the developments, uses, and issues related to Internet testing in 2002. Much to our surprise and disappointment, we found scarcely any published research on Internet testing.

It was with this experience in mind that I decided that this issue of *Testing International* would focus on the theme of Internet testing. Most of the invited brief articles for this issue stem from the ITC's just completed Conference on Computer-Based Testing and the Internet. This theme issue begins with a brief paper based on ITC President Dave Bartram's opening keynote address on the issues, challenges, and opportunities of Internet testing. Next, Tom Oakland discusses ethical implications of computer-based and Internet testing for psychologists. Bruno Zumbo describes how Bayesian statistical methodology can solve some key measurement problems that have emerged in computer-based and Internet testing. Finally, John Kleeman reports on a new British Standard Code of Practice for using information technology to deliver assessments.

There were many excellent presentations at the conference but, given the limited space in *Testing International*, I could only solicit a small few. However, I hope the articles in this issue will prompt discussion and I strongly encourage the presenters from the recent ITC conference and others involved in Internet testing to consider submitting their papers to *Testing International*, the *International Journal of Testing*, and other appropriate outlets so we may benefit from their work.

In the meantime, I hope you enjoy this selection of papers and learn something new about current work in the field of Internet testing.

## PRESIDENT'S LETTER

Prof. Dave Bartram  
 SHL Group  
**U.K.**

Dear ITC Members:

It is difficult to believe that my two-year period of service as President is now nearly at an end. The time seems to have flown. At the start of my term (in July 2000), I identified four areas as priorities.

1. ITC involvement in, and sponsorship of, conferences and symposia on testing issues.
2. A project to develop guidelines for the use of computer-based tests. In particular, to develop guidelines covering good practice in the use of the Internet for testing.
3. Development of the ITC website as a key information and advice centre about tests and testing.
4. Work with the European Federation of Psychologists Associations (EFPA) on the development of test review criteria and exploration of the possibility of coordinating a technical 'benchmarking' review process for tests.

It will be useful to look back and see how far we have progressed on these goals. The first two of these areas have become closely related through the ITC Conference in Winchester.

### 1. Involvement In and Sponsorship of Conferences and Symposia

The major task here has been the organisation of the ITC Conference in Winchester (see below for an update).

I have also spoken at a number of conferences and other meetings in the past six months on the ITC's role and the work on our various Guidelines projects. These include presentations in Sweden, Germany, UK, and USA.

### 2. Guidelines on Computer-based Testing

I have been working closely with Dr Iain Coyne (Hull University, England), who was co-opted onto the ITC Council at the end of last year. Iain and I will be reporting on some work we have done looking at Internet test administration issues at the ITC Conference in Winchester. We will also take responsibility for taking the outcomes from the Conference forward as the basis for drafting guidelines.

### 3. ITC Website

The ITC website (<http://www.intestcom.org>) has now had over 8000 visitors. In the past six months alone, there have been over 5000 visitors to the site – many of these will have been visiting for information about the ITC Conference. This is the only area of the website that has been regularly updated. It would be good if the ITC could identify the resource necessary for more regular updating and editing of content on the site, especially information on issues, debates in testing and links to other websites.

### 4. Work with EFPA: Test Review Criteria

The EFPA Criteria are now available from the EFPA website: <http://www.efpa.be/>.

These criteria have been welcomed in Europe as an important step towards harmonisation of standards for test evaluation. Their focus is very much on traditional tests as these still dominate the market. However, they also contain an important new section dealing with computer-generated expert reports.

### The International Journal of Testing

The first two issues of Volume 1 of IJT have now been published and the final two issues will appear as a single publication in July. As I noted in my last report, it is a matter of considerable regret that delays have been incurred in getting these issues to our members. However, we are determined to catch up on this so that we will be back on track in terms of timing as Volume 2 comes out.

### The ITC Guidelines on Test Use

The Test Use Guidelines continue to be widely disseminated and used. The official translations that have been approved now include English, Danish, Dutch/Flemish, German, Norwegian, Slovenian, South American Spanish (Argentina), Spanish and Swedish. Other translations are continuing to be carried out.

### The ITC Winchester Conference: June 12-15, 2002

I was delighted to announce in April that we had reached our maximum capacity of 200 delegates and that the conference was full! However, due to pressure of numbers, we re-organised the venue arrangements and increased the numbers to 250. In the end, we had 254 people coming from 21 different countries.

By all accounts, this conference has been a tremendous success. We have had very positive feedback - both unsolicited and more formally through the conference evaluation forms - from those who attended. Not surprisingly, the planning and preparation took more and more time as the June deadline approached, and became a major task for my Team Administrator at SHL, Connie James. The ITC Council is grateful both to her, and to SHL, for making her time available.

The programme was very exciting with a wide range of issues and topics being addressed. Delegates were particularly pleased at the wide range of issues covered and quality of the presentations. We were also delighted by the venue. The Guildhall itself was a very attractive place to hold a conference and nothing was too much trouble for the Guildhall staff. They made the experience unforgettable both for us as organisers and for all the presenters and delegates.

We were also very successful in attracting sponsorship. This is important not only for the financial support this provides, but also because it indicates the serious involvement of the testing industry in the issues we will be discussing at this conference. I'd like to take this opportunity once again to thank all the sponsors, who are:

- SHL Group plc
- Educational Testing Service (ETS)
- The British Psychological Society (BPS)
- The College Board
- CAT\*ASI
- NCS Pearson
- nferNelson
- Riverside Publishing
- Thomson Prometric (incorporating Galton Technologies)
- The European Federation of Psychologists' Associations
- The Psychological Corporation
- Psychological Assessment Resources Inc (PAR)

I am planning to publish the conference abstracts on the website shortly. We are also planning to produce a book based around the keynote presentations and other plenary contributions. The current plan is that this will appear towards the end of next year (2003).

However, the key purpose of the conference was to feed into the development of ITC Guidelines. In that respect, it was a great success and has provided us with plenty of food for thought and

substance for taking this project forward over the next two years.

#### Singapore IAAP Congress, 2002

The IAAP Congress in Singapore (7-14<sup>th</sup> July 2002) will be the venue for the ITC's next General Meeting. Members will already have received notification of this from Jacques Grégoire, our Secretary. The nominations for Council Member and Officers have matched exactly the vacancies, so there will be no need for a vote on elections. However, we still hope you will attend the General Meeting to contribute your thoughts and ideas about what the ITC should be doing.

Please note that the General Meeting date is now scheduled for Wednesday 10<sup>th</sup> July in the afternoon (2.00-4.00). The venue will be announced in the Congress programme.

By the time you read this, the Winchester Conference will be behind us and we will be looking forward to Singapore, welcoming Bruce Bracken into the position of President and another interesting and exciting two years for the ITC and its members.

#### BRIEF ARTICLES

##### **Testing on the Internet: Issues, Challenges and Opportunities**

Dave Bartram  
SHL Group plc  
UK

As the market for Internet-delivered computer-based testing develops, and as the technological sophistication of the products increases, so the issue of ensuring those using such tests and assessment tools follow good practice will increase in importance.

In my presentation, I described what the Internet is and how it is becoming part of the fabric of our everyday lives. As it does, so it will become increasingly the medium of choice for test delivery. As a medium, it has both strengths and weaknesses. In the presentation, I considered 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 Internet – closing the 'digital divide'

Having reviewed and illustrated some of the various ways in which the Internet is being used to support testing and assessment in the work and organisational psychology field (see Bartram, 2000 for more details), I focused on some of the issues we need to consider when developing a structure for good practice guidelines.

In so doing, I thought it important to emphasize the fact that Internet technology provides the opportunity for exercising much greater control over the 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 version(s)
- Control over prior practice - enabling test takers to start from a level playing field
- Control over test users – ensuring that only qualified users can access relevant tests
- Control over test takers – authentication: knowing who is taking the test
- Control over test conditions - ensuring conformity to good practice

Much of the concern over Internet testing relates to issues of good practice in three main areas:

1. Ensuring that there is adequate control over the management of the assessment process.
2. Ensuring that feedback and reporting is of high quality and contained within procedures that reflect good practice in assessment.
3. 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.

#### Assessment supervision functions

In the presentation, I described six main functions associated with traditional test administration:

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; that they understand their role in the process and their rights and responsibilities).
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 (e.g., managing problems with equipment, hardware, disruptions during the test session, test taker disabilities).
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 an instrument administration requires the presence of a human supervisor (or remote monitoring by one) 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.

#### Modes of test administration

Four modes of Internet test administration can be defined.

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 local 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 logon with a username and password.
3. *Supervised Mode*. For this mode, local 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 to logon for 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 local supervision is assumed and there is also control over the test-taking environment (e.g., dedicated test centres). 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.

For each of these modes, a number of assessment scenarios were defined (e.g., career guidance, pre-employment screening). Guidelines need to address good practice issues for each mode of administration in the context of relevant scenarios and taking account of the nature and format of the test and the reasons why testing is taking place. For example, is the test one with right answers or it is a 'typical performance' measure; is it being used in a high stakes situation or not; is there a follow up validation process or not?

#### Other issues

The presentation also touched on the need to define good practice in the provision of feedback and reporting of test results. Most computer-generated test reports are designed for the test user rather than the test taker (Bartram, 1995). Considerable

care and attention needs to be given to reports that are intended to provide the sole source of feedback for the test taker.

Finally, the question of test quality was addressed. 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 (and more especially, speeded) 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).

More recent research, reviewed in the presentation, also showed that when sample differences are carefully controlled, there are no differences between traditional supervised paper-and-pencil administration of a personality inventory (SHL's OPQ32i) and administration of the same inventory online with no local supervision. The scale means, scale reliabilities and scale intercorrelations were the same under both conditions.

### Conclusions

The Internet allows us to exercise far more control than we have been able to do in the past over some important aspects of the distribution of materials, management of the assessment process and the collection of data. However, the Internet has many 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. From the test user and test taker's points of view, it is becoming increasingly difficult to discriminate between good tests and bad tests. 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 in the future.

### References

- Bartram, D. (1994). Computer based assessment. In C.L. Cooper & I.T. Robertson (Eds.) *International Review of Industrial and Organizational Psychology*, 9, 31-69.
- Bartram, D. (1995). The role of computer-based test interpretation (CBTI) in occupational assessment. *International Journal of Selection and Assessment*, 3,

178-185.

Bartram, D. (2000). Internet recruitment and selection: Kissing frogs to find princes. *International Journal of Selection and Assessment*, 8, 261-274.

---

### **Ten Ethical Issues Associated with Computer-based Testing and the Internet**

Thomas Oakland  
University of Florida  
USA

The impact of computer-based testing technology in many western countries is enormous. Ironically, its impact as a computer generally is minimal while its impact as a word processor is considerable. This technology has had its greatest impact on promoting communication, not computation.

Technology is developed, introduced, and reviewed within a social context. All people are entitled to access it and derive benefits from its use. A society allows technology to be introduced when it believes its use will improve life by promoting efficiency, increasing safety, reducing expenses, enhancing the attainment of personal and social goals, and extending the quality of life. The attainment of these goals requires attention to ethical issues to help insure technology is used to serve people well.

1. Psychologists do not discriminate in the provision of services on the basis of gender, age, social class, or disability category.

Computer-based (and thus Internet) test technology is unlikely to be used with a representative segment of the world's population and thus has the potential to promote bias. For example, this technology is likely to be utilized by non-handicapped middle class educated males less than 50 years of age who read and write one of three or four languages.

More boys than girls receive an education and acquire facility with computers. Those who have not attended school within the last ten or more years are unlikely to be computer literate. Persons with disabilities are likely to benefit from this technology only after methods that help them accommodate to this technology are implemented. Thus, efforts to insure the technology's widespread availability are needed.

2. Psychologists use assessment methods in a manner appropriate to an individual's language preference and competence.

Two language components are considered when reviewing possible language bias: language dominance (i.e., the language one uses best to communicate) and language competence (i.e., whether one's listening, reading, speaking, and writing abilities are properly developed).

Computer- and Internet-based tests currently use few of the world's 200-plus languages and dialects. Thus, persons whose language differs from English, French, Spanish, or a few other widely used languages are unlikely to be able to utilize this technology and thus will be deprived of its benefits.

Computer-based tests generally require persons to be able to read and write. These two language processes are less likely to be developed than one's ability to listen and speak. Persons with diminished ability to read or write also are likely to be deprived of the benefits of this technology.

3. Psychologists insure sufficient information is available to form opinions.

Psychologists make decisions based on information from multiple methods and sources, that reflects multiple traits, and provides information on behaviors in multiple settings. In contrast, information from computer-based tests may not meet this standard. Its assessment methods and sources generally are limited, and the range of behaviors assessed typically is narrow. Thus, information from this technology may be insufficient to form valid opinions.

4. Psychologists insure privacy and promote confidentiality.

Psychologists take reasonable precautions to insure personal privacy and protect confidentiality. However, no organization or professional can insure personal privacy and confidentiality of data transmitted electronically. All systems are potentially accessible to others. Furthermore, data may be sent to incorrect recipients. Psychologists offering computer- and Internet-based technology services should inform users of the risks to privacy and limitations on confidentiality.

5. Psychologists create and work to maintain, disseminate, store, retain, and dispose of records in a responsible fashion.

Most computer- and Internet-based assessments are intended to be used with large numbers of persons. Issues associated with the accurate recording, dissemination, storage, retention, and disposal of data often are complex and troublesome, and the probability of error increases as data sets increase in size. Clerks who lack sufficient training as well as a commitment to ethical issues often maintain data files. Psychologists offering computer-based technology services should inform users of related risks.

6. Psychologists use assessment instruments in a manner and for purposes that are appropriate in light of research on or evidence of the usefulness and proper application the techniques.

7. When working with diverse populations, psychologists use appropriate assessment instruments whose validity and reliability have been established with that population. When such instruments are not available, care is taken to interpret test results cautiously, with regard to the potential bias and misuse of such tests.

The above two issues are somewhat similar and thus are discussed together. Technical properties of computer-based tests should be held to standards similar to those for paper-and-pencil tests (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999). Data that help describe a test's convergent and discriminant validity are provided in a format that promotes their dissemination and understanding. Moreover, data that examine test validity in light of possible age, gender, and race-ethnicity differences should be made available to test users. Tests that lack these qualities should be used with caution.

8. Psychologists assure the accuracy of the person taking a test.

The administration of most tests requires examinees to produce two or more forms of evidence, including at least one that contains a picture, to verify their identify. Similar assurances are needed for computer-assisted tests. Lacking such evidence, one cannot guarantee the accuracy of the identify of persons completing a test.

9. Psychologists maintain test security.

The maintenance of security of computer-based tests represents a formidable challenge. The ability to print and in other ways download tests as well as

to transmit them to others limits their security. One cannot assume test takers are seeing a test for the first time. A test no longer is standardized when some test-takers have the advantage of seeing some or all test items before taking the test. Tests that lack proper security should not be used.

#### 10. Psychologists and the organizations with which they work provide evidence as to the accuracy of scoring and reporting systems.

One assumes a computer program designed by a company to batch score and report test data provides reliable information. Numerous examples, however, can be provided that invalidate this assumption. Errors exist in such programs, leading to miscalculations and thus misreporting of data, resulting in errors in decisions made on the basis of these data. Methods that both examine the accuracy of data scoring systems and report errors when they occur are needed.

#### References

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). Standards for educational and psychological testing. Washington, DC: American Psychological Association
- Oakland, T. (in press). Selected ethical issues relevant to test adaptation. In Hambleton, R., Spielberger, C., & Meranda, P. (Eds.). Test adaptation. Mahwah, NY: Erlbaum Press.
- (This latter reference can be requested through an e-mail to me at: [oakland@coe.ufl.edu](mailto:oakland@coe.ufl.edu).)

### **Contemporary Uses of Computer- and Internet-Based Testing, a Particularly Appealing Context for the Application of Bayesian Statistical Methods**

Bruno D. Zumbo  
University of British Columbia  
**CANADA**

There are two particularly thorny measurement problems endemic to contemporary uses of computer- and Internet-based testing that find appealing solutions in Bayesian statistical methodology: (a) estimating item parameters in the context of small sample sizes, and (b) computing and reporting test scores that arise from measurement opportunities (i.e., items or tasks) that include binary and polytomously scored items in the same test - and perhaps items that are linked

to a common stem or context (sometimes called "testlets"). The purpose of this short essay is to highlight that Bayesian statistical methods have appealing properties to help solve these two measurement problems.

With an eye toward the purpose of this essay, I will provide a bit more detail as to why these two measurement problems are commonly found in computer-based (and thus Internet) testing and then I will say a few words as to what the essential feature of Bayesian statistical methodology is that helps us resolve these two measurement problems. For a detailed discussion and review of the application of Bayesian methods in measurement, I recommend a forthcoming paper by Rupp, Dey, and Zumbo (2002).

#### Why do these measurement problems arise in contemporary computer-based testing?

What follows are three essential features of contemporary computer-based testing:

1. Many contemporary computer-based tests are not fully adaptive, per se, but rather involve sequential selection of items or sets of items that are formed "on-the-fly". These sets of items are sometimes called "shadow tests" or "forms" and may be composed ahead of time according to the overall test information (i.e., psychometric statistical information which is related to the standard error of measurement) and item content constraints.
2. The essential feature described in #1 above necessitates large item pools that need calibration but, by design to maintain test security, the items in this pool cannot be exposed often. What this translates to is that one needs to estimate the item parameters for the items in this pool in the context of small sample sizes.
3. Today's test developers are not constrained by orthodox thinking that dictates that tests should be comprised of disjoint items that are all scored the same way – i.e., that they are all binary items or, more generally, that they all have the same number of points per item. Instead, today's test developers think in terms of "measurement opportunities" rather than items or tasks. These measurement opportunities can be some combination of binary and polytomously scored events and not necessarily disjoint events. For example, a test taker who wants to be certified as an accountant may be given a computer-based test to assess their accounting knowledge or skills. The test taker is

presented with a series of 10 accounting “problems” that each involve as few as 5 measurement opportunities whereas other “problems” involve as many as 15 measurement opportunities. Furthermore, some of these measurement opportunities are scored in binary format whereas others are scored polytomously.

Taken together, these essential features of contemporary computer-based testing give today’s test developer a great deal of flexibility as to test format. Because these tests are more fluid in form and isomorphic to behaviours and tasks that individuals experience in their daily lives, these contemporary tests may allow for more valid inferences from the test scores. Unfortunately, much of psychometric theory has been developed with the prototypical test being a set of disjoint items that are each scored with the same number of points (e.g., a test made up of all binary disjoint items). Along with their advantages, the new test formats have brought the two measurement problems described above for contemporary psychometric modeling.

### Some Basics of Bayesian Statistical Methods

A brief non-technical overview of Bayesian methodology will help illuminate how this methodology can help us in this context. Bayesian methodology has a fascinating history and, until recently, has not been met with the warmest reception among researchers, philosophers, statisticians, and probability theorists alike. In fact, it is one of those important scientific ideas that was discovered and re-discovered several times over. The methodology is named after the 18<sup>th</sup> century Presbyterian minister, Thomas Bayes, who did important philosophic work in what we would now call scientific methodology and the philosophical foundations of probability. Later, the mathematician Laplace did work that went a long way to show researchers how Bayes’ ideas could be applied. In the social and behavioural sciences, Bayesian methods often have been discussed in the context of decision theory and statistical hypothesis testing (see, for example, Zumbo and Hubley’s 1998 discussion of statistical power).

One of the key features of Bayesian methods is described in the notion that probability statements (and, in turn, point estimates of parameters and corresponding confidence bounds on these estimates) are proportional to the relative conditional probabilities (i.e., the likelihood) times the prior probability. That is, imagine two random

variables X and Y. Now further imagine that you had information about the conditional probability,  $\Pr(x|y)$ , which tells you what to believe about X if you knew the value of Y. Finally, given that you learn that X equals some value x, what do you know about Y? This probability,  $\Pr(y|x)$ , is proportional to  $\Pr(x|y)$  times  $\Pr(y)$ .

$$\Pr(y|x) \propto \Pr(x|y) \times \Pr(y)$$

This, in words, translates to: “The posterior probability” is proportional to “The Likelihood”  $\times$  “The Prior Probability”.

The blending of the likelihood and the prior information is the key advantage of Bayesian methods. It is the mechanical aspects of this blending that has kept it, until relatively recently, from everyday use in statistics and psychometrics. Advances in computers and statistical theory have made Bayesian methods computationally feasible. Conceptually, of course, Bayesian methods allow us to factor expertise and prior knowledge into our computations.

### How can the Bayesian framework help computer-based testing specialists?

The Bayesian framework can help us reduce the sample size needed for item parameter estimation in an item response theory (IRT) model. The recent ITC conference on computer-based testing and the Internet in Winchester, England included some research based at the University of Massachusetts – Amherst (U.S.A.) and ETS in Princeton (U.S.A.) that demonstrated the utility of Bayesian methods in reducing sample size requirements for calibrating items with smaller sample sizes by incorporating prior information about item parameters obtained from item writers or other experts. Ongoing research with my colleagues and graduate students explores the small sample characteristics of Bayesian IRT item parameter estimates as well as robust Bayesian methods and their use with complex survey data. In essence, by using prior information in the process of estimation, one appears to be able to estimate item parameters well with smaller sample sizes than commonly considered with (2- and 3-parameter) IRT models. See Rupp, Dey, and Zumbo (2002) for a description of the very fruitful strategy for sampling from the posterior distribution with computer-intensive methods such as Markov Chain Monte Carlo (MCMC) simulation.

Like the matter of sample size for parameter estimation, the matter of mixed item types and testlets has been thoroughly treated by a unified Bayesian modeling framework introduced by Wainer and his colleagues Bradlow, Wang, and Du (see Wang, Bradlow, & Wainer (2002) for a review of their approach). Their framework allows for the modeling of richer data structures by writing an expanded IRT model and then using Bayesian modeling strategies to solve what, until recently, was an unsolvable statistical estimation problem.

In fact, it is this last application of Bayesian methodology that highlights one key strength of Bayesian approaches. In a fairly straightforward manner, one can now solve very complex statistical modeling problems with data that can be more readily found in measurement and research contexts. It is this advantage that contemporary computer-based testing experts are exploiting.

#### Endnote

<sup>1</sup> I first came across the term “measurement opportunities” in its current use by Professor Ric Luecht and his computer-based testing collaborators at the AICPA in implementing Luecht’s CAST model for computer-based testing.

#### References

Rupp, A., Dey, D. K., & Zumbo, B. D. (2002). To Bayes or not to Bayes, from whether to when: Applications of Bayesian methodology to item response modeling. Forthcoming paper, University of British Columbia, Vancouver, B.C., Canada.

Wang, X., Bradlow, E.T., & Wainer, H. (2002). A general Bayesian model for testlets: Theory and applications. Applied Psychological Measurement, 26, 109-128.

Zumbo, B. D., & Hubley, A. M. (1998). A note on misconceptions concerning prospective and retrospective power. Journal of the Royal Statistical Society, Series D (The Statistician), 47, 385-388.

Author's Note: In the field of Bayesian statistics, people talk about the varieties of Bayesians that can be found: philosophical Bayesians, subjectivist Bayesians, etc. This short essay is yet another step for this author in outing himself as a “closet Bayesian”. The author is Professor of Measurement, Evaluation, and Research Methodology (MERM), and a member of the Department of Statistics and Department of

Psychology at UBC. Correspondence regarding this essay should be addressed to the author at [bruno.zumbo@ubc.ca](mailto:bruno.zumbo@ubc.ca).

#### **BS 7988: A New British Standard on using IT in Delivering Assessments**

John Kleeman  
Questionmark  
UK

The British Standards Institution ([www.bsi-global.com](http://www.bsi-global.com)) has recently set up a committee called IST/43 to participate in ISO SC36 international standards in learning, education and training and to consider British Standards in this area. IST/43 has members from government, universities, companies in the industry and exam boards.

One of IST/43’s first projects has been to draft and publish a new British Standard Code of Practice on using information technology (IT) to deliver assessments, called BS 7988. This standard was developed by a subcommittee of IST/43 chaired by the author, and the standard was published in April 2002. With the widespread and increasing use of computerized testing, BSI felt that it was important to set out guidelines to ensure that candidates are fairly assessed and that people can rely on the results of computer assessments.

BS 7988 is not a mandatory standard to follow – it is a code of practice that organizations can choose to follow or not. It has had wide debate and obtained consensus within the UK assessment community, prior to release. BS 7988 only covers the delivery of assessments, not their preparation or post-delivery analysis; these are important, but less easy to standardize on due to differing practices. Due to the scope of IST/43 (and of the international committee ISO SC36), BS 7988 formally only covers assessments in education and training, and not those used in recruitment.

The standard sets out how organizations should organize their delivery of assessments to ensure that students can take tests fairly and securely. Some clauses are strict, for example:

- For assessments longer than 1.5 hours and where the candidate works almost entirely at the screen, there should be provision for the candidate to take a break.

Some clauses are vaguer, for example:

- If the candidate is permitted to quit the assessment before the expiry of the time limit, consideration should be given to asking the candidate to confirm that the submission of the answers is final, particularly if some items remain unanswered.

And some clauses require procedures to be set up depending on the particular needs of the assessments and users:

- Assessment distributors should provide assessment centres with clear instructions on how to load and configure assessment software (including any associated software for administration) and how to begin and end assessments.

The standard also includes some guidance on accessibility – giving guidance to organizations delivering exams on how to make reasonable adjustments for candidates with disabilities, so that no candidate is placed at a substantial disadvantage. This is particularly important given recent legislation in the UK (and in some other countries) on this subject.

BS 7988 is available at a moderate charge from BSI. A data sheet including price and how to purchase is at <http://edd.bsi.org.uk/link.php3/ist/43>. Although most of the Standard is likely to be useful internationally, parts of it do reference UK legislation. A longer term possibility is to consider adapting it for international use and moving it to ISO or European standardization.

It is also possible in the future that BSI or other organizations may set up some sort of compliance system, to allow organizations who follow BS 7988 to receive some sort of formal confirmation that they are compliant.

BS 7988 has just been launched, and how useful and significant BS 7988 is will depend on its adoption in the community. However, since the consequences of failures in assessment delivery can be very serious, both for the candidate and for the organization delivering the assessment, it is likely that assessment providers will want to show that they are following all good practice that they can.

The purpose of BS 7988 is to ensure fairness and build confidence in computerized assessments – I would commend it to anyone interested in this subject.

## THE NOTICEBOARD

### Barbara Byrne Receives Jacob Cohen Distinguished Teaching/Mentoring Award

We are pleased to announce that Prof. Em. Barbara Byrne, School of Psychology, University of Ottawa, Canada and ITC Treasurer, will receive the Jacob Cohen Distinguished Teaching/Mentoring Award by Division 5 (Evaluation, Measurement, Statistics) of the American Psychological Association. Dr. Byrne is recognized for her excellence in classroom teaching, mentorship of students, popular workshops, best-selling books, and numerous articles. Those of you who attended Dr. Byrne's workshop on structural equation modeling (SEM) at the ITC conference in Winchester will have seen her outstanding teaching ability in action. Dr. Byrne's research focuses on measurement issues related to self-concept, burnout, and depression as well as the sound application of structural equation modeling in the test validation process. As noted in the December issue of *Testing International*, Dr. Byrne will also be receiving the Distinguished Contributions to Education and Training Award given by APA this August.

## UPCOMING CONFERENCES ON TESTING & ASSESSMENT

### 30<sup>th</sup> International Congress on Assessment Center Methods

October 15 -18, 2002

Pittsburgh, Pennsylvania, United States

Website:

<http://www.assessmentcenters.org/pages/intlconference.html>

Contact name: Cathy Nelson

Contact e-mail: [Cathy.Nelson@ddiworld.com](mailto:Cathy.Nelson@ddiworld.com)

The International Congress on Assessment Center Methods, sponsored by Development Dimensions International is an annual conference for assessment and HR professionals to discuss trends, research and new developments.

**International Conference on Questionnaire Development, Evaluation, and Testing Methods**  
 November 14-17, 2002  
 Charleston, South Carolina, United States

Website: <http://www.jpsm.umd.edu/qdet>  
 Contact name: Jennifer Rothgeb  
 Contact e-mail: [jennifer.m.rothgeb@census.gov](mailto:jennifer.m.rothgeb@census.gov)

The conference will focus on research that develops, applies, and/or evaluates new methods of questionnaire development. It is also concerned with the special requirements and challenges of testing questionnaires for special populations, including children, diverse cultural and language groups, and organizations. The conference will focus on interviewer-administered and self-administered questionnaires. Based on the conference papers, a monograph presenting state of the art research findings will be produced through Wiley Publishers.

**24th Annual Language Testing Research Colloquium**  
 December 12-15, 2002  
 Hong Kong

Website:  
<http://www.engl.polyu.edu.hk/ACLAR/ltrc.htm>  
 Contact e-mail: [egACLAR@polyu.edu.hk](mailto:egACLAR@polyu.edu.hk)

The theme this year is Language Testing in Global Contexts. This conference takes place immediately before the 13th World Congress of Applied Linguistics (AILA 2002) in Singapore.

**Technology in Testing: Application and Innovation**  
 February 24-26, 2003  
 Amelia Island, Florida, United States

Website:  
<http://www.testpublishers.org/Con2003/2003cover.html>

Last year, the Association of Test Publishers (ATP) hosted a conference on Computer Based Testing in California. No further information about this new conference is currently available so keep checking their website if this conference is of interest to you.

**Consortium for School Networking's (CoSN) 8<sup>th</sup> Annual K-12 School Networking Conference: Achievement, Assessment & Accountability**  
 February 25-27, 2003 (with International Research Symposium on Feb. 28)  
 Arlington, Virginia, United States

Website: <http://www.k12schoolnetworking.org/>  
 Contact name: Michelle Shirley  
 Contact e-mail: [michelle@cosn.org](mailto:michelle@cosn.org)

In the wake of *No Child Left Behind*, there is greater emphasis on ensuring that classroom interventions impact achievement and that the decisions of school leaders are data driven. *SUBMISSIONS MUST BE RECEIVED BY SEPTEMBER 13, 2002.*

**7<sup>th</sup> Annual CSU Fullerton Assessment & 2<sup>nd</sup> GE Embedded Assessment Conferences**  
 March 12-14, 2003  
 Fullerton, California, United States

Website: <http://faculty.fullerton.edu/AssessmentConf/>  
 Contact name: JoAnn Carter-Wells  
 Contact e-mail: [jcarterwells@fullerton.edu](mailto:jcarterwells@fullerton.edu)

Assessment: The Compass for Guiding Learning through Currents of Change is the theme for the year 2003 Assessment Conference sponsored by the Cal State Fullerton College of Business and Economics. Rapid changes are taking place in higher education as a result of technology development, changing constituent demographics, and reordering of state and federal government budgeting priorities. This conference seeks to bring experts together to explore the ways that outcome assessment might be used to help institutions of higher learning steer a course through these troubled waters. *SUBMISSIONS MUST BE RECEIVED BY NOVEMBER 4, 2002.*

**Submission deadline for the December 2002 issue of *Testing International* is Nov. 1, 2002.**

**Please submit all articles and reports (preferably as IBM PC-compatible Word or WordPerfect files) to:**

**Dr. Anita Hubley, Editor  
*Testing International*  
 Dept. of ECPS, 2125 Main Mall  
 University of British Columbia  
 Vancouver, BC  
 CANADA, V6T 1Z4**

**Or via e-mail at [anita.hubley@ubc.ca](mailto:anita.hubley@ubc.ca)**

**Please feel free to duplicate copies of *Testing International* and distribute them to your colleagues. Address any correspondence regarding *Testing International* to the Editor at the address shown above.**

**ITC Website**

**<http://www.intestcom.org>**