The British Psychological Society

Guidelines for the Development and use of Computer-Based Assessments

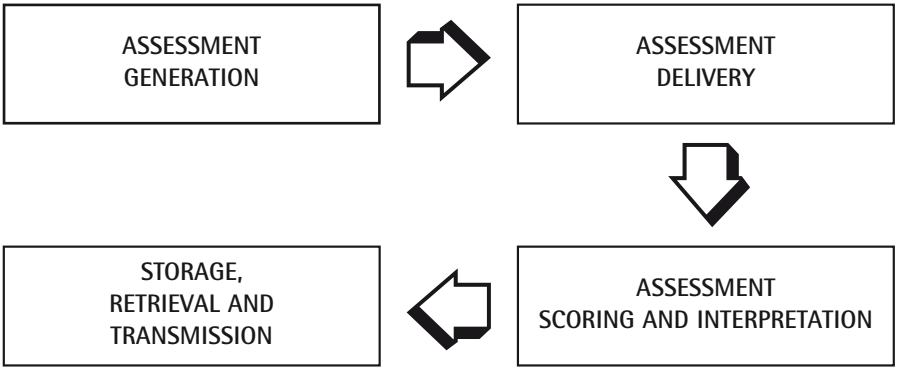
Psychological Testing Centre

www.psychtesting.org.uk

Introduction

Recent years have seen substantial developments in the theory and methods underpinning psychological testing and assessment, as well as growth in the use of psychological testing in employment, educational and clinical assessment. One area that has seen substantial and rapid developments is that of computer-based assessment or CBA.

What is a CBA? Well, while it contains the phrase computer-based, the term CBA may be considered to include any psychological assessment that involves the use of digital technology to collect, process and report the results of that assessment. The figure that follows summarises four components of a CBA system for the delivery of psychological assessment.



Assessment Generation: Traditionally, psychological tests and assessments have been generated by human item writers. Although human item writers can exercise creativity in constructing psychological assessments, given the simple fact of human variability, this process can often result in inaccuracies or errors that are identified in the course of subsequent psychometric analyses. Some items or original assessment content are therefore lost or require further modification. Today, the process of assessment generation can be executed through the use of item engines. These engines encapsulate a model of the intended assessment (say a measure of reasoning or of personality), and take the form of software that writes the test or assessment at the press of a key. The advent of such engines clearly marks an advance in the potential for controlling the quality and cost of the assessment production. More recently, these engines have been enhanced by the use of neural nets (also referred to as artificial intelligence or connectionist engines) that enable the software to learn in the act of production, thereby further enhancing the production of subsequent items and other versions of a test or assessment. However, these item engines have to be validated to ensure that their outputs provide a meaningful basis for assessment.

Assessment Administration: The 1970s and 1980s saw the development of networked computer suites for the administration of psychological assessments, as well as the use of portable personal computers for their administration, scoring and interpretation. The obvious advantages of such computerised administration are greater standardisation and control over the administration process (which should result in gains in the accuracy of assessment), and a potential reduction in assessment costs through the automation of the human labour traditionally involved in administration. More recently, such computerised administration has been enhanced in a number of ways. Among these are adaptive testing of abilities and attainment in which testing is tailored to the individual test taker. The test programme usually begins by administering several items to obtain a first estimate of a person's ability level. Subsequent items varying in difficulty are then administered to refine this estimate until some stopping rule (acceptable level of error in making the estimate) is reached. As well as providing a more efficient (shorter) method of testing, research has shown adaptive testing is also more motivating for the test taker since items which are too easy or too difficult are avoided. Another way in which computerised administration can enhance psychological assessment is through the use of on-line diagnostics. These diagnostics evaluate the test taker's responses and can flag up whether an administration is in some way invalid. It may be that the test taker has not understood the instructions for an ability test, or that answers to a personality test suggest an abnormal pattern of responding. In either case, the assessment may not be accurate and may require further investigation and action. A final example of the enhancements offered by computerised administration is distance assessment. With the advent of intranet and internet, it is now possible for a person at one site to administer an assessment to someone at another physical location that may in fact be in another country. Again, the potential gains in time and access to psychological assessment are considerable.

Assessment Scoring and Interpretation: As well as being time consuming, the scoring and interpretation of psychological tests and assessments may also involve complex calculations (such as in the comparison of scores on different tests or different scores from the same test), and may require substantial expertise in the theory supporting the test or assessment. Computer based test interpretation (CBTI) software packages have been available for some time and are widely used to generate reports on test taker's. More recently, such CBTI packages have begun to incorporate artificial intelligence to provide more sophisticated forms of test score analysis, and to provide narrative reports that appear more natural as if written by a human rather than an automated expert. CBTIs are increasingly used to assist hiring and promotion decisions, and are often extended in their interpretation to likely behaviours at work. Also, with increased concern about the impact of workplace stress on performance and

human resources costs, a rapid area of development is CBTI packages for screening of employees for potential stress disorders.

Storage, Retrieval and Transmission: Computer storage media (e.g. disks, zip drives, CD ROM) provide for a more efficient method of storage, retrieval and data management than paper-based records. With the increasing power of database and communication systems, the gains in cost reduction and flexibility are considerable, not least in the facility to transmit results to clients at remote locations easily and quickly. However, users should familiarise themselves with the Data Protection Act (1998) as well as other Acts of Parliament that relate to the use psychological assessments (e.g. equal opportunities and disability discrimination legislation).

Purpose of these guidelines

UK guidelines and competence standards already exist for the use of psychological tests and for conducting psychological assessments. So, why have guidelines for CBAs? First, while CBAs may sometimes only appear to represent an alternative means of administering psychological assessments, it is important that those responsible for developing and using CBAs are made aware of the issues involved in using this medium of assessment. Secondly, previous CBA guidelines have tended to focus on specific aspects of CBA (e.g. APA, 1986[1], and Green *et al.*, 1984[2]) or have been developed for distribution within specific contexts (e.g. Burke, 1993[3]). Thirdly, while the principles underlying valid test and assessment construction equally apply to CBAs, the ease of CBA construction does pose a serious threat to effective and fair practice in the use of psychological tests. This has been highlighted by commentators in the US and the UK.

> '... the development of computer-based tests and test interpretations has become a cottage industry in the worst sense of the word. For those with even minimal skills, software development is an easy entry business that offers the chance for a large profit on a small investment.'
> Schoenfeldt (1989)[4]

---

[1] American Psychological Association (1986). *Guidelines for computer-based tests and interpretations.* Washington, DC.

[2] Green, B.F., Bock, R.D., Humphreys, L.G., Linn, R.B. & Reckase, M.D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement, 21,* 347–360.

[3] Burke, E. (1993). NATO technical guidelines for the development of computer-based tests. In E. Burke & P. Van Raay (Eds.), *Computer-based Assessment in NATO: Final Report of Research Study Group 15.* AC/243 (Panel 8) TR/12. Brussels: NATO Headquarters.

[4] Schoenfeldt, L.F. (1989). Guidelines for computer-based psychological tests and interpretations. *Computers in Human Behavior, 5,* 13–21.

> '... CBA provides the test developer and user with potentially far greater Assessment power... With that power comes greater responsibility to ensure that the information generated is used wisely ... [CBA] will also pose new problems for professional practice and ethical standards. The professional bodies will need to address these issues. Care is needed that guidelines are produced which will nurture and not stifle potential growth.'
> Bartram (1994)[5]

Building on the British Psychological Society's Level A (ability) and Level B (personality) standards for test use, these guidelines seek to reaffirm the following principles:

Principle 1: That, as with all psychological assessments, users should be made aware of what constitutes best practice in CBA so that they can make informed evaluations and choices between CBA systems offered to them.

Principle 2: That CBAs should be supported by clear documentation of the rationale behind the assessment and the chosen mode of delivery, appropriateness and exclusions for use, and research evidence supporting validity and fairness.

Principle 3: Requirements for administration of the CBA should be clearly documented and should include the knowledge, understanding and skills required for competent administration.

Principle 4: The knowledge, understanding and skills required for interpretation of CBA information and for the provision of such information to a third party should also be clearly stated.

The structure of these guidelines is modelled on the Code of Fair Testing Practices in Education developed by the US joint Committee on Testing Practices[6]. Using the four components of a CBA described in the introduction to these guidelines, they seek to state the responsibilities of the two principal parties in the act of testing: those who develop (Developers) CBAs and those that use them (Users).

Of paramount importance in the relationship between these two parties is clear information provided by developers on the qualities of the CBA being offered for use, and the skills required of a user to interpret this information in order to make judgements suited to the needs of an assessment.

---

[5] Bartram, D. (1994). Computer-based Assessment. In C.L. Cooper & I.T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology*. John Wiley and Sons.

[6] *Code of Fair Testing Practices in Education* (1988). Washington DC. Joint Committee on Testing Practices.

ASSESSMENT GENERATION 

What is assessed?

DEVELOPERS

- Should provide clear documentation of what is being measured, for whom such an assessment is appropriate, and how CBA supports valid meaningful psychological assessment.
- Where the CBA is a computerised version of an assessment previously administered through paper-and-pencil or apparatus means, should provide clear documentation of the equivalence between the CBA and non-CBA versions of the assessment.
- Where algorithms are used to generate items or assessment content (such as item engines), or where algorithms are used to control the order of assessment (as in adaptive testing), the principles underlying the algorithms and the research evidence that they produce reliable, fair and valid assessment should be clearly documented.

USERS

- Should ensure that the knowledge, understanding and skills required to evaluate the appropriateness of the CBA in meeting an assessment need are available and applied.
- Should ensure that documentation describing research on the reliability, validity and fairness of the CBA is reviewed and understood.
- Should be able to provide a clear statement of the objectives of the assessment, why the CBA was selected for use, and what other alternative methods were considered.
- Should evaluate the necessary resources, physical and human, necessary to procure, use and maintain the CBA for ongoing use.

How is the assessment undertaken?

DEVELOPERS

- Should provide a clear description of the hardware and peripheral devices required to administer the CBA.
- Should ensure that:
  - Screen information, text, graphics and auditory information is clearly legible.
  - Screen displays giving information on time, question or task labels, feedback on performance, paging and help facilities are logical and compatible with the purpose of assessment.
  - Interfaces such as keyboards, touch screens and light pens should be logical and compatible with the sequence of actions required to respond to the CBA.
  - Screen displays and interfaces do not disadvantage members of the target population for which the CBA was designed.
- Should provide clear descriptions of the environmental conditions (e.g. work area, lighting) required for administration of the CBA.

USERS

- Should ensure that the knowledge, understanding and skills required to evaluate the hardware, software and environmental requirements of the CBA are reviewed and understood.
- Should ensure that documentation of administrative (hardware, software, environments) requirements are reviewed and understood.
- Should ensure that CBA displays and interfaces do not unnecessarily impede test taker's with disabilities and special needs.
- Should ensure that the knowledge, understanding and skills required for competent administration of the CBA, including procedures for operating CBA equipment and software, are available and applied.
- Should ensure that the knowledge, understanding and skills for managing CBA use and maintaining the CBA for ongoing use are available and applied.

Guidelines Section 3

ASSESSMENT SCORING AND INTERPRETATION

How is the assessment scored and interpreted?

DEVELOPERS

- Should state clearly the principles underlying the scoring and interpretation of the CBA, and the research evidence supporting the validity of scores and interpretations.
- Should state the knowledge, understanding and skills required for competent interpretation of CBA scores.
- Should state clearly the procedures for operating CBA scoring and interpretation software (CBTIs), including any diagnostic information on the validity of a test taker's responses to the CBA.
- Should provide advice on appropriate use of CBA scores in supporting decisions regarding those assessed.

USERS

- Should ensure that the principles underlying the scoring of the CBA, and the research evidence supporting those principles are reviewed and understood.
- Should ensure that the knowledge, understanding and skills required for competent use of the CBA scoring and interpretation software are available and applied.
- Should ensure that the knowledge, skills and understanding required for competent interpretation of CBA scores, including validity checks on test takers scores, and for the provision of score interpretations to a third party are available and applied.
- Should review and ensure the understanding of appropriate use of CBA scores and interpretations in making decisions and recommendations about those assessed.

What procedures are required for storage and use of the asssessment information?

DEVELOPERS

- Should provide clear documentation on facilities incorporated in the CBA system for storage, retrieval and transmission of CBA scores and their interpretations.
- Should clearly document recommendations and procedures for establishing levels of security and access to CBA scores and interpretations.
- Distance assessment may involve the delivery of the CBA to a test taker at a remote site for administration, or the transmission of scores and interpretations to a client at a site other than where the CBA was delivered. Where distance assessment is a feature of the CBA system, procedures and systems for verifying the identity of test taker's or client's, and for maintaining the confidentiality of the assessment should be clearly documented.
- Procedures and systems for maintaining the security of stored CBA data and legal requirements that may apply to such storage should be clearly documented.

USERS

- Should ensure that the storage, retrieval and transmission facilities of the CBA system are reviewed and understood.
- Where distance assessment is to be undertaken, should ensure that procedures and systems for verification and maintaining confidentiality of assessment are known and understood.
- Should ensure that the means of securing access to CBA systems, scores and interpretations are known and understood, and that procedures are in place to minimise unauthorised access to systems and information.
- Should ensure that legal requirements under the 1998 Data Protection Act regarding storage of and access to CBA scores and interpretations are reviewed and understood.

Contributors

The following helped to develop these guidelines.
Professor Dave Bartram, *International R&D Director, University of Hull*
Professor Graham Beaumont, *Royal Hospital for Neuro-disability, Putney*
Professor Mike Berger, *Royal Holloway College, University of London*
Eugene Burke, *Chair, Steering Committee on Test Standards*
Lisa Cramp, *Saville and Holdsworth Ltd*
Dr Robert Feltham, *PA Consultants Ltd*
Wendy Lord, *ASE, NFER-Nelson*
Stuart Robertson, *Stuart Robertson and Associates*
Dr John Rust, *Goldsmiths College, University of London*
Dr Chris Singleton, *University of Hull*
Patrick Tapsfield, *Human Assessment Laboratory, University of Plymouth*

The following also contributed to two symposia sponsored by the Steering Committee on Test Standards to test both interest and opinion in guidelines for CBAs.

Occupational Psychology Conference, University of Warwick, January 1995.
Professor Dave Bartram
Eugene Burke *(Symposium Chair)*
Wendy Lord
Dr Colin Selby, *Selby Millsmith Ltd*

Occupational Psychology Conference, Blackpool, January 1997.
Professor Dave Bartram
Eugene Burke *(Symposium Chair)*
Dr Robert McHenry, *Oxford Psychologists Press*
Dr David Wright, *University of Plymouth*

British Psychological Society Test User Conference, Brighton, June 1998.
Eugene Burke *(Symposium Chair)*
Lisa Cramp
Dr John Rust
Dr David Wright

Prepared by the Professional Practice Board Steering Committee on Test Standards
Reviewed April 2002

The British Psychological Society
The Psychological Testing Centre, St Andrews House, 48 Princess Road East, Leicester LE1 7DR.
Telephone 0116 252 9530 Facsimile 0116 247 0787 E-mail enquiry@psychtesting.org.uk
Website www.psychtesting.org.uk

PTC05/10.05