ELSEVIER

Archives of Clinical Neuropsychology 22S (2007) S49–S61

Archives of CLINICAL NEUROPSYCHOLOGY

Development and quality assurance of computer-based assessment batteries

Robert E. Schlegel, Kirby Gilliland

Center for the Study of Human Operator Performance (C-SHOP), The University of Oklahoma, 3200 Marshall Avenue, Suite 260, Norman, OK 73072, United States

Abstract

The purpose of this article is to outline critical elements in the development and quality assurance (QA) assessment of a computer-based assessment battery (CAB). The first section of the article provides an overview of the life cycle of a representative CAB, typical evolutionary stages, and many of the essential considerations for designing and developing a CAB. The second section of the article presents a model for conducting a quality assurance assessment of a CAB. A general narrative of several steps in the QA process is supported by a table of recommended QA assessment elements. Although this QA process model may not be definitive for all cases, it provides a general framework within which a systematic assessment of any CAB can be conducted.

Keywords: Quality assurance; Computer-based assessment battery; Test battery

1. Development and quality assurance of computer-based assessment batteries

The introduction of microcomputers in the 1970s soon led to the development of computer-based versions of classical paper-and-pencil and electromechanical tests of human capabilities. These computer-based human performance tests soon evolved into more advanced computer-based assessment batteries (CABs) that provided automated and systematic assessment of a broad range of skills and abilities (Gilliland & Schlegel, 1993; Kane & Kay, 1992; Perez, Masline, Ramsey, & Urban, 1987; Schlegel, Hockey, & Gilliland, 2003; Shingledecker, 1984). These CABs quickly provided demonstrated effectiveness in human performance research and real-world applications including personnel screening and selection, general medical applications in the areas of neuropsychology and disease assessment, and a vast array of educational, industrial, and business applications. At the same time, many CABs possessed irksome problems that compromised their usefulness, such as test design inadequacies, inadequate instructions for installation and use, poor or no instructions for those individuals being tested, data filing and management inadequacies, and programming errors. The fact that these issues survive to this day, even in relatively sophisticated CABs, often results from advancing the CAB into use before it has received even the most basic levels of a systematic quality assurance (QA) evaluation. The cost of such haste can be catastrophic in terms of faulty tests, compromised data, and wasted time, effort, and resources.

Over the past two decades, the authors have been intimately involved in CAB research. Their studies have involved (1) the use of CABs to investigate environmental and person variables that moderate human performance, (2) the relative comparison of different CABs, (3) the internal structure and integrity of CABs, (4) the psychometric properties of CABs, (5) the training requirements for CABs, and (6) the establishment of normative data for CABs. The authors

E-mail address: schlegel@ou.edu (R.E. Schlegel).

0887-6177/$ – see front matter
doi:10.1016/j.acn.2006.10.005

have also been asked by many CAB developers to review their batteries for strengths and possible areas of concern. In addition, they have been frequently contacted by researchers wanting to know such information as which version of a specific test is the most reliable or valid, how one version of a test compares to a similar version in another CAB, and other practical questions such as how many practice trials are needed to train a person to asymptotic performance levels. The authors are currently assuming responsibility for the future development and enhancement of the Automated Neuropsychological Assessment Metrics (ANAM[1]), one of the most successful and comprehensive automated neuropsychological test instruments evolving from a long line of Department of Defense CABs. Through these experiences it has become abundantly clear that a systematic and detailed quality assurance review is crucial to the development of any CAB.

The purpose of this article is to review some of the critical steps in CAB development and CAB quality assurance assessment with the intent of highlighting areas of interest and concern for CAB designers and users. The discussion is intended to ultimately improve the development, evaluation, maintenance, and application of CABs. Although this treatment is not necessarily complete or definitive, this article offers a way of opening an important dialogue within the scientific and user community of CABs. Due to the broad scope of such an endeavour, many of the statements in this document are generalizations based on the authors' two-decade history of work in computer-based assessment. While exceptions to the generalizations may be plentiful, these observations are intended to demonstrate the numerous benefits associated with the application of even a modest level of QA in CAB development or *post hoc* evaluation.

This article has two broad sections. The first section provides an overview of the life cycle of a representative CAB, typical evolutionary stages, and many of the essential considerations for designing and developing a CAB. An examination of the evolutionary stages reveals several opportunities to enhance CAB quality and features and to reduce potential problems *during* the development process. Not all CABs undergo every stage, and some CABs may undergo additional stages, but the consideration of a general model of CAB development is useful in many ways. The second section of the article presents a model for conducting a quality assurance assessment of a CAB. Although typically implemented *post hoc*, the basic principles could be profitably applied earlier in the development process. A general narrative of several steps in the QA process is supported by a table of recommended QA assessment elements. Although this QA process model may not be definitive for all cases, it provides a general framework within which a systematic assessment of any CAB can be conducted.

2. The life cycle of a computer-based assessment test/battery[2]

As software products, computer-based assessment tests and batteries (i.e., collections of tests) cannot escape one of the cornerstones of software engineering—the software life cycle. This cycle has been observed several times in the history of CABs, often driven by technology advances such as the platform migration from DEC VAX and PDP minicomputers, to Apple II and Commodore 64 microcomputers, to personal microcomputers running DOS, then Windows, and more recently to Personal Digital Assistants. One representation of the software life cycle incorporates the following activities:

- Origins of the CAB: Theory-Based or Problem-Based.
- General Test Requirements and Specifications.
- Detailed Software Development and Battery Testing and Evaluation.
- Psychometric Properties, Validity, and Sensitivity.
- Factors Affecting Test Battery Utilization.
- Funding for Development and Support.
- Marketing, Promotion, and Sales of the CAB and Support Services.

[1] ANAM is a registered trademark.

[2] In many cases, single computer-based tests are developed when a selected human skill, ability, or function is evaluated or is to be used for screening. In fact, each test within a multi-test battery could be viewed in this manner. We will not confound the discussion with additional acronyms (e.g., CAT for computer-based assessment test). The basic tenets concerning test life cycle and QA can easily be applied to single-test or multi-test implementations.