Assessing Language through Computer Technology

Reviewed by Daniel Waller, University of Central Lancashire, UK

Computer-Assisted Language Testing (CALT) is definitely on the rise. Major examinations such as TOEFL, Pearson PTA and Password use computer adaptive testing (CAT) and boast high levels of reliability as well as allowing for increasingly easy access to their tests for students around the world. Cambridge now uses a computer-based platform for its writing markers, allowing for greater levels of reliability in scoring and a far more manageable system. Elsewhere, studies are being carried out on such areas as eye-tracking on screens when candidates take examinations, while other projects conduct corpus-based analysis of different levels of learner English or the use of multi-word chunks in natural language. This type of utilization of technology for the purposes of specification of language performance, the nature of competences, and quality assurance is becoming, if it is not already, the industry standard. And furthermore, much of it is accessible. With the average laptop able to make the computers on the original Starship Enterprise look a bit weedy, and the easy availability of software and education platforms, CALT is open to most of us to use and evaluate. With all of these developments it is essential that teacher training, whether pre or in service, provides language professions with knowledge of what CALT is, how it differs from traditional language testing and what theoretical principles underpin its development and how it should be evaluated.

This book is a potential point of entry for such matters. It is intended for those “professionals who work to help the learners who use CALT” (p. 2). Chapter one begins by considering the issues raised by CALT for three groups: teachers, test developers and language testing researchers and provides the reader with an overview of many of the areas and issues which will be examined in the book while giving a range of examples of CALT in use. These range from the provision of feedback to learners on their performance on individual items in a test

DOI: 10.4018/jvple.2011070106
through to discussions of the implications for the use of multimedia in tests. Perhaps most interesting is the discussion of how CALT raises the potential of being able to assess new constructs such as time taken on each item or multiple facets of performance on a cloze test – where correct or erroneous spelling, syntactic fit, derivational and morphological performance could be simultaneously measured allowing for five measures from a single item. Indeed, it is in discussions like this where this book is at its best, examining some of the unique opportunities that CALT offers and this is very much the focus of the following chapter.

Chapter two moves on to examine how CALT differs from ‘traditional’ testing methods and explores some of the opportunities that it presents in more detail, such as the possibility of having a wider range of input sources such as streamed video, audio and images in tests, and the reliability of computer marking. Throughout the book, the presentation of examples to illustrate these points bring home that what Chapelle and Douglas are talking about is already happening. As well as setting out many of the opportunities that CALT presents, chapter two gives a nod to some of the potential challenges and these are explored in more detail in chapter three.

Six key challenges are set out in chapter three, which the writers detail and examine what research has been carried out with regard to these issues. The difficulties presented include the issues around how a computer-based assessment might lead to a different level of performance from a paper-based (or otherwise traditionally administered) test, issues related to the use of adaptive item selection by a computer and inaccurate automatic response scoring. Other issues which are discussed are certainly not limited to computer-based testing; factors such as the use of new task types, the issues of security around tests and the potential negative washback of tests have all been discussed extensively in conventional testing, but as the writers identify the considerations shift somewhat in a CALT context. However, despite these challenges the writers state that “based on existing research, we were not able to find any strong evidence suggesting that these threats are so insurmountable that CALT should be inherently suspect than other types of test” (p. 60). In fact, one of the real strengths of the book is not only the critical examination of existing research but the calls for further examination of such issues. Seen from this extent, CALT can be seen to be raising the game for all assessment.

Chapter four will be of great interest to those teachers and test designers who want to dip their toes in CALT. The first half of the chapter is an illustration of how the authors have used WebCT to develop a test for their own context. Readers are also pointed towards software such as Hot Potatoes (Half-Baked, 2004) and the Discovery School Test Centre (Discovery School, 2004) which provide easy-to-use test writing platforms which teachers can utilize for their own tests. Such software offers teachers a ‘rough and ready’ way of producing computer-based tests at relative speed, but as the authors point out, these systems lack the functionality required for more sophisticated CALT. A four-process architecture for an assessment system devised by Almond, Steinberg and Mislevy (2004) is presented as a tool for the evaluation of authoring tools. The system is itself based on Chapelle’s (2001) suggestion that such a system would need to meet the wide and various requirements of different purposes of assessment, including language assessment. Almond, Steinberg and Mislevy’s system identifies the need for such a design to be able to carry out processes such as estimating the level of task difficulty, analyzing learners’ use of English, creating and organizing the objects used in constructing test and instruction, gather process data and provide a structure for collecting relevant data in a learner model.

While the previous chapters consider CALT from the point-of-view of the non-specialist, chapter five, “Evaluating CALT,” takes up the issue of how such tests should be appraised from the perspectives of experts in the field of computer-based assessment. What may surprise those new to the area is how long there have been proposals for how to carry out such
evaluation; Chapelle and Douglas cite papers from 1984 and 1988 which initially set out to address some of the issues which are unique to CALT. Most focus in such evaluations is on the actual user interface, the work of Noijons (1994) and Fulcher (2003) demonstrate that questions of layout, user-friendliness and clarity are overriding concerns for many evaluations. Particularly useful in the chapter are signposts to research into the issues outlined in the third chapter which will provide those interested with key further reading. Chapter five ends with an examination of the overall usefulness of CALT using Bachman and Palmer’s (1996) model with a discussion on reliability, construct validity, authenticity and interactiveness. The book assumes familiarity with these terms, so those without a background in assessment will need to go and look at the original, but the discussion remains accessible and clear throughout.

The final chapter, “The Impact of CALT,” examines the current and potential role that CALT could play in a range of areas including the language classroom as well as SLA studies. The authors suggest that CALT has not yet had as substantial an impact as some predicted a decade or more ago. It is implied that this has less to do with the technology available than with the clear specification of language and language use. Projects such as the English Profile (www.englishprofile.org), which is an attempt to specify the language that constitutes the levels of the Common European Framework of Reference for Languages (CEFR), is an example of the sort of research that might enable the revolution which Chapelle and Douglas suggest is waiting in the wings. The final chapter, like chapter five, contains useful references to research that those interested in the wider picture of CALT may find interesting.

There is no doubt that this is a very accessible and useful book for those new to the area or who have been using programs such as WebCT or Hot Potatoes and wish to extend their practical and theoretical knowledge of CALT. It assumes a reasonable level of familiarity with the technology, but in the view of this reviewer this is appropriate to the audience of the book. Given the limited publications in the field to date, the focus on CALT frameworks and the “how to” approach, coupled with the signposting to further reading, means that the book still remains relevant for students and researchers. Like Computer-Assisted Language Learning, CALT is subject to constant changes. Nevertheless, the approach taken in this book will ensure that it remains essential reading for those coming into or seeking to widen their knowledge of the field.

REFERENCES


Daniel Waller has been involved in ELT since 1993 and taught for nine years in Turkey. He’s been responsible for a many different projects in previous posts, including developing beginner level English programmes, business English classes, and word lists and revising a text book for students preparing to enter university. He completed the Diploma in English Language Teaching to Adults in 2000, and an MSc in Teaching English for Specific Purposes in 2005. Daniel is currently working on a PhD and researching the role of discourse in defining performance in timed writing examinations at levels B2 and C1 of the Common European Framework of Reference for Languages. At the University of Central Lancashire he is heavily involved in the development of English language examinations and his main research interests are in the areas of testing, written discourse and lexis.