

COMPUTER-BASED TESTING: BACKGROUND, ISSUES, AND NEW DIRECTIONS FOR ASSESSMENT

E. Marcia Johnson and John W. Brine
Department of General & Applied Linguistics
The University of Waikato, Hamilton

The term 'computer-based testing' (CBT) means using a computer to present test questions (items) and collect and score responses. There are numerous types of CBTs and their development and use represent differing views of assessment. For language teachers and administrators, the array of possibilities as regards CBT can sometimes be confusing, if not overwhelming. This paper will describe and discuss some of the different perspectives on CBT in order to provide a background to the field and guidance about possibilities for using CBT in the classroom. Descriptions of item-banking, computer-adaptive testing (CAT), and web-based testing (WBT) will be provided, including discussion of the advantages and disadvantages of each. This will be followed in the second half of the paper by an exploration of uses of computers in alternative assessment, with practical classroom-based examples being provided. The paper concludes with a discussion of future directions for CBT.

Introduction

As the name implies, computer-based tests are administered and scored using a computer. This means that not only is instant feedback possible for the test-taker (including diagnostic feedback), but also questions (items) can be tailored to individual test-taker's proficiency levels. Typically, item types include multiple-choice, true/false, and matching as these have been the easiest to generate and score on a computer, and responses are gathered through point-and-click or drag-and-drop techniques. More recently, 'authentic' forms of test items, including role play, interviews, and presentations have become possible using multimedia, CD-ROMs, or interactive video. Web-based testing, a relative 'newcomer' to CBT, has also become increasingly popular during the past few years. Some computer-based tests, including the Test of Written English (TWE), a component of the Test of English as a Foreign Language ((TOEFL) – an internationally-accepted English language proficiency test), permit use of word processing for essay writing. Thus, in a very broad definition of CBT, word processing could be included, but it will not form part of this more focused discussion.

This paper will describe briefly such aspects of CBT as item-banking, computer-adaptive testing (CAT), and web-based testing (WBT). Advantages and disadvantages of each within second language instruction will be explored, followed by a general discussion of alternative assessment

approaches that are both classroom-based and computer-supported. The paper will end with an exploration of future directions for CBT.

Early CBT and item banking

In the area of language assessment, CBT has been used since the PLATO project (University of Illinois) of the 1960s, but was not adopted more generally in education until about ten years later (Godwin-Jones, 2001). According to Brown (1997), the term 'item banking' refers to "any procedures that are used to create, pilot, analyze, store, manage, and select test items so that multiple test forms can be created from subsets of the total "bank" of items" (p. 44). Since the 1970s, extensive research into the design and administration of item banks has been conducted by measurement specialists and has focused primarily on the development of computer-adaptive testing (Chalhoub-Deville, 2001).

Within item banking, individual test items, predominantly of the multiple-choice, true/false, or matching type, are developed, organised, and stored in databases, originally on mainframe computers, but now also on microcomputers. Test makers can specify the types of items and levels of difficulty they require, and different test forms can be generated from the large range of available stored items. Microcomputer based item banking and test creation software provide teachers with increased flexibility to design items and produce their own tests although the software can be expensive, depending on its level of sophistication. (See Godwin-Jones (2001), and LeLoup and Ponterio (2001) for more comprehensive descriptions of available online testing resources and microcomputer-based testing software.)

Although item banking has been used extensively to produce assessments across a broad range of school subjects, it was not particularly popular with language teachers in the years following 1980, until recently. The main reason for this has been the emphasis within communicative language teaching on performance assessment, requiring that students actually produce examples of written or spoken language (Chalhoub-Deville, 2001). Such items have been difficult to implement and score within CBT as answers are often multi-faceted and require the judgment of human raters. Although the situation is changing, many CBTs still remain narrowly focused on determining whether or not the test-taker has mastered discrete aspects of language structure and form (Chapelle, 2001). Such items, due to their dichotomous 'correct' or 'incorrect' nature, can be scored objectively by a computer with relative ease.

In spite of the limited range of item types in many item banks, CBT has a large and rapidly growing role in 'high-stakes' language tests (high stakes tests are those for which the results may hold significant consequences for the test-taker). For example, the Educational Testing Service (ETS) in Princeton, New Jersey is responsible for the TOEFL, which has been offered in computerised form since 1998. As of 2001, all paper-based versions of the test were scheduled to be discontinued (Alderson, 2000; Brown, 1997), however, this goal has not yet been achieved (Educational Testing Service, 2001b). The University of Cambridge Local Examinations Syndicate (UCLES) is producing CD-ROM versions of some of its tests and is also in the process of implementing a computer-based version of the International English Language Testing System (IELTS). In this new version, the test will be offered at selected testing centers, but as an optional alternative to the traditional paper-based form, not as a replacement (as has occurred with the TOEFL).

Arguably, one of the main advantages of CBT is that items can be computer-adaptive; however, one of the main disadvantages focuses on a possible lack of construct validity due to an emphasis on a restricted subset of possible item formats. Our discussion will first consider aspects of computer adaptive testing (CAT) within a 'traditional' testing framework before turning to a description of alternative assessments in which computers can play an important role.

Computer-Adaptive Tests (CAT) within the CBT framework

In a CAT, test items are drawn from an item bank and each test form is tailored to the individual test-taker's level of proficiency as determined by responses made during the test. A schematic representation of computer-adaptive testing can be referenced in Figure 1. At the beginning of a test, the test-taker is presented with items of 'medium' difficulty. If the items are answered correctly ('C' on Figure 1), slightly more difficult items are presented, however, if the first items are answered incorrectly ('IC' on Figure 1), the next ones are slightly easier. Thus, the computer updates the estimate of the test-taker's proficiency level after each item has been answered and then uses that estimate to determine the selection of subsequent items (Norris, 2001; Rudner, 1998). In the computerised form of the TOEFL, both the listening and structure sections of the test are computer-adaptive.

Building an Adaptive Test

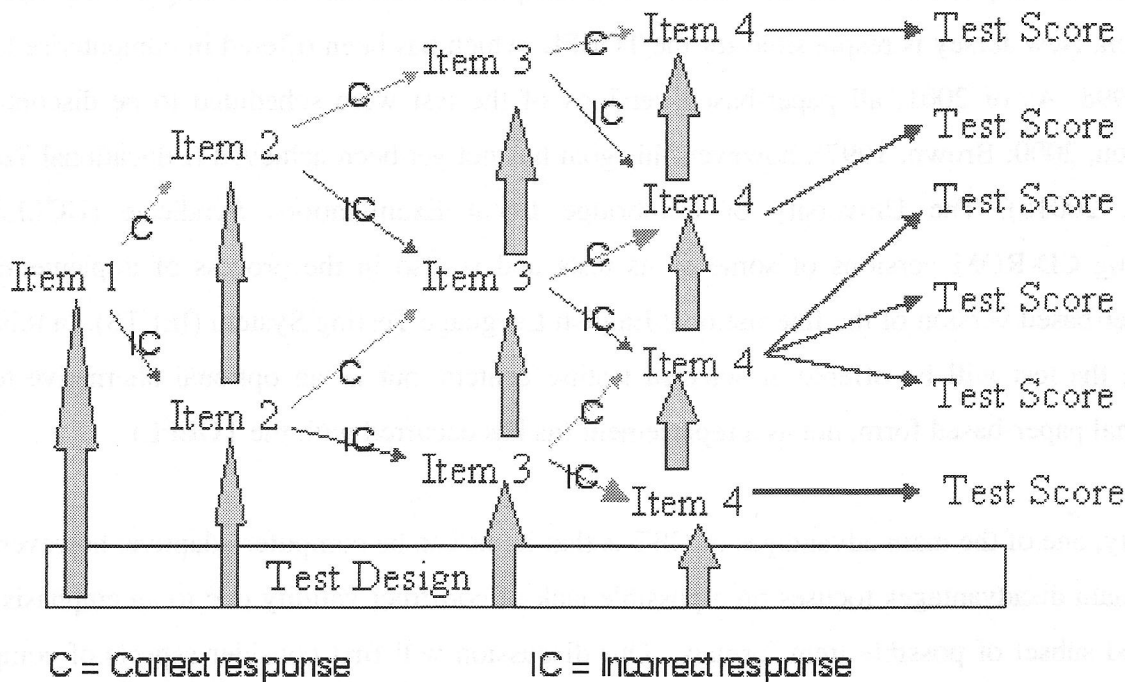


Figure 1: Schematic of computer-adaptive testing (Rudner, 1998).

Advantages of CAT

Efficiency and adaptability

One of the clear advantages of CAT is that test-takers are seldom presented with very easy or very difficult items for their proficiency level. Whereas in a traditional paper-based test, test-takers typically encounter all levels of questions – some of which are much too easy or much too difficult for them – in the CAT, only items that are close to the test-taker's level are presented. One advantage of this is that accurate assessments can be completed much more quickly as test-takers are answering questions that are more appropriate for their level. Once a stable estimate of the test-taker's proficiency has been established, the test ends. Brown (1997) estimates that a CAT can accurately assess a test-taker's proficiency level in about 45 minutes as opposed to the usual 2 or 2.5 hours required for similar paper-based tests. Also, because the test is adapting to the estimated level of each individual test-taker, the assessment is highly personalised. Dunkel (1997) states that even if a group of as many as 50 test-takers is assessed at the same time, it is unlikely that any two people will receive exactly the same test, provided, of course that the item bank is large enough and sufficiently developed to include a wide range of items (Dunkel, 1999).

Security

In the case of high-stakes proficiency testing, much of which relies on computer-adaptive items, CBT offers greater security and flexibility. For example, the logistical problems of shipping paper-based tests to remote testing centres, the possibility of shipments of completed tests being lost, or the necessity of testing only on pre-arranged dates can all be eliminated in a CBT (Roever, 2001).

Disadvantages of CAT and CBT

Scoring

Some of the controversial aspects of CAT are concerned with how items are scored. Whereas in a paper-based test, subsequent questions may prompt a test-taker to reflect, reconsider, then change, an earlier answer, this is not possible in most CAT environments. The CAT, sensitive as it is to each individual response, cannot support later corrections since the response to each item affects the selection and presentation of subsequent items. Also, for obvious reasons, item omissions can present serious problems, and consequently, in most CAT software, the test-taker *must* supply an answer in order to proceed through the test.

Another problematic aspect of a CAT is determining the cut score, that number which determines when the test-taker's level of proficiency has been reached. Brown and Iwashita (1996) in an investigation of a CAT placement test in Japanese, reported that students from different L1 backgrounds (English and Chinese) made significantly different types of errors on their test. As Brown and Iwashita (1996) made clear, there is ample research evidence to support L1 interference in acquisition of linguistic forms.

When a CAT model of item difficulty is developed, it is based on actual test results from large numbers of candidates who have taken paper-based versions of the test. The model of item difficulty thus derived is then built into the CAT. However, if the model is based on the types of errors made by students from one particular L1 background, but students from a different L1 background actually sit the test, then many students may not fit the model. In fact, this was precisely what Brown and Iwashita (1996) found in their study. Such a problem has implications not only for the setting of cut scores, but also for the ethics of test development, particularly when tests will be administered across different cultural groups. Much of the research into CAT design has focused on addressing these concerns (Dunkel, 1999).

Computer familiarity

Another key issue in a discussion of disadvantages of CBT involves the possible introduction of non-construct relevant bias into assessments if test-takers are not familiar with using a computer (Brown, 1997; Russell & Haney, 1997). Fulcher (1999) in his study of the use of CBT in a (high-stakes) English language placement test found that there may indeed be a computer skills-related bias, particularly for test-takers from certain L1 backgrounds. Kirsch et al (1999), also found that test-takers' varying familiarity with computers affected their test performance; however, Roever (2001) found that browser familiarity had no significant effect on test performance. Various high-stakes testing agencies claim that even test-takers with no prior computer experience can learn how to take a CBT by following a short tutorial before the test begins (Educational Testing Service, 2001a; Graduate Management Admission Council, 2000). The important point here is that lack of computer familiarity can be addressed through tutorials and that in this regard, classroom teachers could play a role through use of 'low-stakes' CBT. For example, web-based testing (WBT) could be a beneficial tool for preparing students for high-stakes tests.

Item development

Beyond issues of lack of computer familiarity, an essential point in a discussion of CAT and CBT is that item development is very complicated, and assumes a sound knowledge of item-response theory (Chalhoub-Deville & Deville, 1999; Dunkel, 1999; Roever, 2001). Developing a valid and reliable CAT does not occur through chance or experimentation. It is for this reason that other forms of CBT, including web-based testing (WBT) and alternative computer-based assessments, might prove much more practical and accessible for classroom teachers.

Web-based testing (WBT)

Web-based testing, unlike other forms of CBT, does not require expensive standalone software; tests are written in HyperText Markup Language (HTML) and stored on the tester's server. Versions of the test are then downloaded to the test-taker's computer, but the scores are collected and processed on the server. Results can either be displayed on the test-taker's computer at the end of the test or sent to the tester's email address. Item formats are flexible, although are frequently of the multiple-choice, true/false, or matching type, as is common across many CBTs. Since the test is web-based, rather than being resident on a particular computer, test-takers can be provided with the Uniform Resource Locator (URL) of the testing website and then take the test 'anytime – anywhere' (Fulcher, 1999; Roever, 2001).

Although tests are written in HTML, there are numerous free, web-based test editing tools to assist the teacher-developer, for example, test authoring sites such as the *QUIA Corporation* (<http://www.quia.com/>), *Hot Potatoes* (<http://web.uvic.ca/hrd/halfbaked/>), or the *Virtual Language Centre* at the University of Hong Kong (<http://vlc.polyu.edu.hk/Authoring/authorin.htm>) are quite straightforward for teachers to access and use. Knowledge of HTML is really not required in order to create simple web-based tests, and free space for storing them is readily available on the web (Roever, 2001).

‘Ready-made’ WBT

Even though it is not difficult to create web-based tests, it can be time-consuming. For teachers who want the convenience of web-based testing, but feel they do not have the time to create their own tests, there are numerous online sites at which ready-made web-based quizzes and tests can be found. For example, quizzes in a variety of languages can be tried at *Dave’s ESL Café* (<http://www.eslcafe.com/>), *Self-study Quizzes for ESL Students* (<http://www.aitech.ac.jp/~iteslj/quizzes/>), *The ESL Study Hall* (<http://gwis2.circ.gwu.edu/~gwvcusas/>), or *Teaching with the Web* (<http://polyglot.lss.wisc.edu/lss/lang/teach.html>), to name but a few. Often it is assumed at these sites that students will be working on their own through the tests or quizzes.

Advantages of WBT

Efficiency and adaptability

Without doubt the biggest advantages of WBT are the low cost and the ease with which tests can be created by non-specialist teachers. Roever (2001), in his recent discussion of WBT, describes this as “testing goes grassroots” (p. 86). Because of the ease with which tests can be created and administered, teachers have a great deal of flexibility to develop a range of assessments and specifically tailor them to support and reinforce ongoing classroom instruction (Ellis & Tebbutt, 2001).

For ‘ready-made’ quizzes, there is no test creation required at all on the part of the teacher. As long as teachers and students have access to computers and an Internet connection, they can find potentially thousands of easy-to-use language quizzes or tests simply by typing the required URL.

Disadvantages of WBT

Security

WBTs can be also be created as computer-adaptive tests (Fulcher, 1999) although much more complex programming and item development are required, and consequently, this type of WBT tends to be of the high-stakes variety. Roever (2001) describes several disadvantages of WBTs and all of these pertain to high-stakes testing. Foremost is the problem of guaranteeing test security in a web-based environment; for example, knowledgeable computer 'hackers' could break into a web-based item bank and compromise items. In addition, transferring sound or video files, large as they tend to be, can be slow and is not always reliable. For these reasons, Roever (2001) concludes that WBT is not a good candidate for high-stakes testing, yet Fulcher (1999) in his description of WBT for university language placement tests avoids the issue of test security entirely.

Restricted test formats

For ready-made WBT, there are also many distinct disadvantages. Often quizzes follow predictable, repetitive formats so that students visiting a particular site may become bored with the testing style. For example, many ready-made WBT tend to focus on testing discrete points of grammar and are predominantly text-based. As a result these quizzes may appeal to lower / intermediate level students, but there are few sites at which more advanced students can find tests appropriate for them. Not all sites are maintained on a regular basis, and so a student may return to what seemed to be an interesting site, only to find that nothing new has been added since the last visit. Another disadvantage of ready-made WBT is that quiz content may bear no relationship whatsoever to material the teacher is actually using in the classroom.

Uneven quality of websites

Last, but not least, the process of identifying and evaluating high quality websites is extremely time-consuming; many sites are simply not worth a visit. Thus for the busy teacher, although it may initially seem a 'time-saver' to use ready-made WBT, in the long run, it may actually be much more efficient to create one's own WBT materials and tailor them to content being taught in the particular instructional context.

Discussion

What seems clear from the preceding descriptions is that there are two distinctly different approaches to CBT – one focusing on high stakes testing, and to which few teachers or students will

ever have meaningful input (Hamp-Lyons, 2000), and one which is potentially much more accessible to a wider circle of people having a stake in language assessment (Roever, 2001). Certainly not all assessment is of the high-stakes, gate-keeping variety.

Towards formative CBT assessment

Over the past 25 years, there has been a shift in assessment from being primarily summative in purpose to being more formative. This has meant that in all educational contexts, including language teaching, assessment has moved from being a formal process detached from pedagogical considerations to a procedure which is integrated into teaching and learning. It is widely recognized that the learning *process* is at least as important as the learning *outcomes* (Genesee & Upshur, 1996). However, as has been discussed, the focus of many CBTs and CATs, in spite of their potential to be otherwise, remains largely summative. Chapelle (2001, p. 131) in her discussion of computer adaptive language testing states that many of the questions raised by Canale in 1986 about how computers can improve both the theory and practice of assessment have yet to be answered. Such questions include, for example, the role of computers to support alternative assessments reflecting the multi-dimensional, communicative nature of language.

Learner-adaptive language testing

Nevertheless, selected examples of alternative approaches to summative CBT do exist. Kane-Iturrioz (1997) discusses a formative approach to assessment, *learner-adaptive* language testing, in which the learner can adapt the test to him or herself rather than the other way around as in a CAT. In learner-adaptive language testing, items are stored in a computerised database (as in item-banking), but students use software to select the types of test items with which they feel most comfortable or areas in which they want reinforcement. In this manner control over the types of questions the student will answer rests with the individual and is based on personal reflection as regards perceived language needs and item type preferences. The student has access to 'help' facilities such as an 'explanation' feature into which the teacher can type elaborations of language points, a dictionary, glossary, or 'clues' about what must be considered in order to answer the question. Kane-Iturrioz (1997) also states that students themselves could use the software to create test questions, perhaps working in teams to design and trial tests with each other. This approach is highly consistent with both current learner-centred views of assessment (Ekbatani, 2000) and the idea that 'noticing' plays an important role in language acquisition (Ellis, 1997).

Kane-Iturriz (1997), citing research of Alderson and Wendeatt (1991) states that when students were at liberty to self-select test questions, item types such as multiple-choice and jumbled text were most popular in both paper-based and computerised formats. However, students thought that the computer-based tests were more useful because of the help facility, and they believed that they had learned more vocabulary and grammar because of their control over test content and the process of assessment. A key point about such an approach is that assessment is broadened from serving just institutional selection and placement needs to serving individual learning needs.

Web-based conferencing for continuous assessment

Another important development in the area of alternative assessment is the use of web-based conferencing software to support student peer assessed work (Johnson, in press 2002). In this approach, a web-based computer conference is the focal point for continuous assessment tasks involving reading, writing, and peer-based evaluations. Rather, than the computer being merely a device to present test items, it serves instead as a communicative tool through which students co-construct knowledge and understanding of written texts (Kern, 2000). The purpose of the peer-based assessment is to provide on-going opportunities for writing, reading, and reflecting upon multiple examples of authentic language as the learners interact in electronic space. Although not a 'traditional' testing situation, it is nevertheless a computer-based environment in which meaningful assessment occurs.

Future developments in CBT

What lies ahead for CBT? One of the main criticisms of CBT relates to a perceived lack of construct validity. As has been described, many of the commonly used item types are restricted to the multiple-choice or matching variety, primarily because performance assessments usually involve complex judgments on the part of a human rater. The situation is slowly beginning to change, however, particularly in the areas of writing and speaking as new, expanded item formats and testing approaches are developed (Alderson, 2000).

CBT for assessing writing

A promising new development in the area of assessing student-produced written language, particularly in high-stakes tests, is 'e-rater', a software program developed by ETS (Alderson, 2000; Burstein & Marcu, 2000). The e-rater software is 'trained' by having previously scored examples of test-takers' open-ended writing input to it and through the use of natural language processing

techniques, the software ‘learns’ to duplicate the work of human raters (Alderson, 2000, p. 601). In addition, Burstein and Marcu (2000) report that they foresee future applications for e-rater such as the provision of diagnostic feedback to test-takers on how to improve their writing.

CBT for assessing speaking

Larson (2000) describes oral testing software (OTS) developed at Brigham Young University, which leads teachers step-by-step through the test creation process. Types of prompts or elicitation techniques, student response preparation time, and the response time itself can all be specified by the teacher, and other digital material support materials can be selected for inclusion in the test (Larson, 2000, p. 57).

Another important assessment tool for oral language is PhonePass™ (Alderson, 2000; Bernstein, 1999), which uses speech recognition technology and the telephone. PhonePass™ sample items and test papers are purchased in advance and delivered to a test administrator. Before the actual test, the test taker can use the sample items to become familiar with the testing format and can also consult other instructional resources, including a dictionary or even their teacher, in order to prepare. When ready, the test taker places a call to the PhonePass™ system, reads from the test form, and responds to the spoken instructions – all over the telephone. Upon completion of the test, responses are parsed by a computer, compared to native speaker models, and scores are reported according to repeat accuracy, pronunciation, and fluency (Bernstein, 1999, Ordinate Corporation, n.d.).

Large-scale diagnostic testing

DIALANG, another large-scale CBT, is a web-based diagnostic assessment instrument being developed as part of a European Union funded research project. This system will provide free, internet-delivered proficiency tests for reading, writing, listening, grammatical structures, and vocabulary in fourteen European languages (Anderson, 2000; DIALANG, 2001). Individual items may include both sound files and images. Unlike the TOEFL and IELTS tests, however, DIALANG is not high-stakes, but rather is a diagnostic tool designed to help people assess their proficiency level and receive tailored feedback on how to improve. It is described as being particularly appropriate for people who may have learned another language outside of formal study (DIALANG, 2001). The system is currently being pilot-tested and should be fully operational by the end of 2001.

Alternative assessment

Finally, for classroom teachers, use of WBT (both teacher developed and 'ready-made'), learner-adaptive tests, and web-based communication software for continuous assessment are all promising new directions. Through these tools, assessment can become much more closely integrated into instruction and be used to help students assume more control of their own learning.

Conclusion

Although to date, much of the work in CBT has been summative and associated with high stakes testing, there are numerous developments that can support innovative learner-centred, classroom-based assessments. Learner-adaptive tests can provide diagnostic feedback to both instructors and students, which can lead to improved teaching and learning environments. Students can assume more control over aspects of language acquisition that suit both what they want to learn and what they need to learn. In computer-adaptive and learner-adaptive tests student frustration levels could be reduced as they work, at their own pace, on items more suited to their level of proficiency. In web-based assessment environments students can support and enrich each other's learning through carefully constructed, communicative, peer-assessment tasks. In addition, with improvements to technology, more and varied aspects of language can be assessed thus improving the construct validity of CBT, whether classroom-based or of the high-stakes variety (Bennett, 1998, 2001).

The main concern for teachers, in designing and using CBT is how to integrate it into classroom processes. As in all instructional environments, consideration of such key issues as the main purpose for the assessment and consequences of its use must be foremost in the mind of the developer. Regardless of the testing format, the same basic approach to test development is required – that of defining the language constructs to be assessed (Bachman and Palmer, 1996). Issues of how to design and implement continuous assessment, motivate students, and determine individual learning styles need to be framed, as in any good teaching environment, within the larger context of syllabus design and individual lesson planning.

Assessment is a complex process, yet one in which all teachers engage. A deeper understanding of the role of computers and possibilities for their use in language assessment can provide exciting new directions and possibilities for all educators.

References

- Alderson, C. (2000). Technology in testing: The present and the future. *System*, 28(4), 593-603.
- Bachman, L. F., & Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bennett, R. E. (1998). Reinventing assessment: Speculations on the future of large-scale educational testing. *Educational Testing Service*, 1-19. Policy Information Centre, Princeton, N.J.: Educational Testing Service.
- Bennett, R. E. (2001). How the internet will help large-scale assessment reinvent itself. *Educational Policy Analysis Archives*, 9(5). Available: <http://epaa.asu.edu/epaa/v9n5.html> [Verified 15 October 2001]
- Bernstein, J. (1999). PhonePass™ Testing: Structure and construct. Ordinate Corporation Technical Report. Available: <http://www.ordinate.com/pdf/StructureAndConstruct990826.pdf> [Verified 21 November 2001]
- Brown, A., & Iwashita, N. (1996). Language background and item difficulty: The development of a computer-adaptive test of Japanese. *System*, 24(2), 199-206.
- Brown, J. D. (1997). Computers in language testing: Present research and some future directions. *Language Learning & Technology*, 1(1), 44-59. Available: <http://llt.msu.edu/vol1num1/default.html> [Verified 15 October 2001]
- Burstein, J., & Marcu, D. (16-18 October, 2000). Toward using text summarization for essay-based feedback. Paper presented at the Conférence TALN 2000, Lausanne, Switzerland. Available: <http://www.ets.org/research/talnsum.doc> [Verified 15 October 2001]
- Chalhoub-Deville, M. (2001). Language testing and technology: Past and future. *Language Learning & Technology*, 5(2), 95-98. Available: <http://llt.msu.edu/vol5num2/default.html> [Verified 15 October 2001]
- Chalhoub-Deville, M., & Deville, C. (1999). Computer-adaptive testing in second language contexts. *Annual Review of Applied Linguistics*, 19, 273-299.
- Chapelle, C. A. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research*. Cambridge: Cambridge University Press.
- DIALANG organization. (2001). *DIALANG – Project and system*. Available: <http://www.dialang.org/english/summary.html> [Verified 13 November 2001]
- Dunkel, P. A. (1997). Computer-adaptive testing of listening comprehension: A blueprint for CAT development. *The Language Teacher Online*, 21(10) [On-line serial]. Available: <http://langue.hyper.chubu.ac.jp/jalt/pub/tlt/97/oct/dunkel.html> [Verified 14 November 2001]
- Dunkel, P. A. (1999). Considerations in developing or using second/foreign language proficiency computer-adaptive tests. *Language Learning & Technology*, 2(2), 77-93. Available: <http://llt.msu.edu/vol2num2/default.html> [Verified 15 October 2001]
- Educational Testing Service (2001a). *About the computer-based TOEFL test*. Available <http://www.toefl.org/educator/edabtcbtfl.html> [Verified 15 October 2001]
- Educational Testing Service (2001b). *TOEFL testing program*. Available: <http://www.toefl.org/tstgprog01.cfm> [Verified 15 October 2001]
- Ekbatani, G. (2000). Moving toward learner-directed assessment. In G. Ekbatani & H. Pierson (Eds.), *Learner-directed assessment in ESL* (pp. 1-12). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Ellis, R. (1997). *SLA research and language teaching*. Oxford: Oxford University Press.
- Ellis, Rachel, & Tebbutt, C. (17 August, 2001). *The creation and implementation of on-line learning materials*. Department of General & Applied Linguistics Seminar Series, The University of Waikato, Hamilton, New Zealand.
- Fulcher, G. (1999). Computerizing an English language placement test. *ELT Journal*, 53(4), 289-299.

- Genesee, F., & Upshur, J. A. (1996). *Classroom-based evaluations in second language education*. Cambridge: Cambridge University Press.
- Godwin-Jones, B. (2001). Language testing tools and technologies. *Language Learning & Technology*, 5(2), 8-12. Available: <http://llt.msu.edu/vol5num2/default.html> [Verified 15 October 2001]
- Graduate Management Admission Council. (2000). GMAT test taking strategies. Available: http://www.gmac.com/gmat/preparation/test_strategies.shtml [Verified 5 October 2001].
- Hamp-Lyons, L. (2000). Social, professional and individual responsibility in language testing. *System*, 28(4), 579-591.
- Johnson, E. M. (in press, 2002). The role of computer-supported discussion for language teacher education: What do the students say? *CALICO Journal*, 20(1).
- Kane-Iturriz, R. (1997). Computer-based language assessment: A formative approach. *ReCALL Journal*, 9(1), 15-21.
- Kern, R. (2000). *Literacy and language teaching*. Oxford: Oxford University Press.
- Kirsch, I., Jamieson, J., Taylor, C., Eignor, D. (1998). Computer familiarity among TOEFL examinees. (TOEFL Research Report No. 59). Princeton, NJ: Educational Testing Service.
- Larson, J. (2000). Testing oral language skills via the computer. *CALICO Journal*, 18(1), 53-66.
- LeLoup, J. W., & Ponterio, R. (2001). On the Net: Language testing resources. *Language Learning & Technology*, 5(2), 4-7. Available: <http://llt.msu.edu/vol5num2/default.html> [Verified 15 October 2001]
- Norris, J. M. (2001). Review of computerized adaptive testing: A primer (second edition). *Language Learning & Technology*, 5(2), 23-27. Available: <http://llt.msu.edu/vol5num2/default.html> [Verified 15 October 2001]
- Ordinate Corporation (n.d.). Product description. Menlo Park: California. Available: <http://www.ordinate.com/> [Verified 21 November 2001]
- Roever, C. (2001). Web-based language testing. *Language Learning & Technology*, 5(2), 84-94. Available: <http://llt.msu.edu/vol5num2/default.html> [Verified 15 October 2001]
- Rudner, L. M. (1998). An on-line, interactive, computer adaptive testing mini-tutorial, 11/98. Available: <http://ericae.net/scripts/cat/catdemo.htm#this> [Verified 15 October 2001]
- Russell, M., & Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Education Policy Analysis Archives*, 5(3). Available: <http://olam.ed.asu.edu/epaa/v5n3.html> [Verified 15 October 2001]