

*Online Assessment
of Speaking and Writing:
a guide to current and
emerging technologies*

*Dr. Gary Woodill
Chief Learning Officer
Operitel Corporation*

Online Assessment of Speaking and Writing: a guide to current and emerging technologies

A working paper developed for the
Centre for Canadian Language Benchmarks (CCLB)

Gary Woodill, Ed.D.
Chief Learning Officer
Operitel Corporation

© 2004 by the Centre for Canadian Language Benchmarks
and Operitel Corporation

Introduction

Online assessment of language skills is based on a set of emerging technologies, and is therefore in the midst of rapid change. While it is generally agreed that the online assessment of reading and listening is technically possible today, there is less agreement on whether online assessment of speaking and writing is feasible. This is because it is far easier for computer technology to assess interpretation skills (reading and listening) than to evaluate presentation skills (speaking and writing). Interpretation is about the comprehension of meaning, while presentation is a combination of the creation of meaning and performance abilities.

This paper is a brief description of current and emerging technologies for the online assessment of speaking and writing. Its purpose is to guide further research by identifying and describing new technologies for assessment of speaking and writing, highlighting any commercial products that are now on the market that would accomplish all or part of this task, and listing publications of researchers who are working in this field, particularly in Canada, who could contribute to further discussion and analysis.

There are many issues in the field of assessment as a whole, in the use of online assessment tools, and in the assessment of language skills. It is beyond the scope of this paper to discuss these issues in detail. Instead, the focus is on identifying promising technologies rather than general assessment issues. Because of rapid developments in networked technology, most assessment techniques that were used on standalone computers two or three years ago can now be delivered online. This is important to note, because a review of the literature often offers opinions that are out of date, even if they were published a short time ago.

It is important to distinguish between assessment tools, assessment tasks and the technological platforms for delivering online testing. Thus, the CCLB's requirements that "tasks be culturally accessible to all adults", and that "tasks should be contextualized", need to be reflected in assessment delivery platforms, as well as in assessment tasks themselves. Because both the context and the cultural requirements of assessment tasks can vary greatly, the CCLB may wish to consider a set of "adaptive" online tasks that can be easily changed in terms of context and content. Such adaptive tasks will need to be delivered by adaptive online software that is flexible enough to evaluate the individual users who sign on and give tasks that have been individualized for that person's language, culture, and personal characteristics (e.g., gender or age), while maintaining consistency that allows comparisons across cultural and contextual differences.

It is also important also to realize that language skills such as speaking and writing have a number of sub-skills. Online assessment tools may be able to evaluate some sub-skills and not others. In practical terms, this may mean that a comprehensive evaluation of speaking and writing may be only undertaken today through combination of online testing and non-computerized evaluation. It is important, therefore, to know the purpose of online assessment of speaking and writing, before judging the appropriateness of a particular tool. The tools needed for a thorough assessment will be quite different from those used for basic screening.

Almost 15 years ago, Bunderson, Inouye, and Olsen (1989) identified four generations of computer assessment. These were 1) computerized testing, 2) computerized adaptive testing, 3) continuous measurement, and 4) intelligent measurement. This was a vision of how computerized assessment would develop in the future. There are lots of examples of the first-generation of computerized assessment, while the second-generation is now in the process of being commercially developed. The last two

generations, while being worked on in research labs, are not ready for general use. It is important to recognize where we are in this technology development curve, and to realize that rapid progress is being made. What is not feasible today may very well be available in three to five years.

Computer assessment of language skills covers a wide range of assessment procedures and technologies. Computer Aided Assessment (CAA) includes such activities as on-screen marking, networked peer assessment, and the e-mailing of performance samples for offline evaluation. Atkinson and Davies (2004) list the following as some examples of CAA:

- interactive exercises and tests completed on a computer
- use of computers to produce coursework
- on-screen marking of students' word-processed writing
- use of revision software
- using a spreadsheet or database to keep a record of students' marks
- use of email to send coursework and to receive marks and feedback
- use of Web pages to set tasks for students and to provide tutor support
- use of plagiarism detection software

In contrast with Computer Aided Assessment, Online Computer Assessment (OCA) uses the World Wide Web as a platform for delivering and performing assessment tasks and giving and receiving direct feedback on these tasks. In terms of the assessment of speaking and writing, OCA is a very new technology, with several commercial products being announced only in the past 6 months. Given the history of software development,

rapid improvements in OCA technology for speaking and writing can be expected in the next one or two years, as firms compete for market share.

Components of Online Language Performance Assessment

Part of what makes online assessment of speaking and writing possible today are the advances in the hardware for the production and delivery of online media. High quality yet inexpensive headsets with microphones (\$30-40) allow crisp clear speech, and new compression CODECs (compression/decompression software) allow this to be transmitted efficiently over the Internet. One company even has a test of speech using the telephone as the input device. Digital video cameras that can be mounted on the front of a computer monitor are now less than \$100. The combination of these can produce performance samples that are quite adequate for online transmission and evaluation. And, while many people world-wide may not have access to computer technology, the fact is that it has become inexpensive enough to be installed in public institutions such as libraries or embassies.

The issue of the technology experience of potential test-takers is another matter, however, as this will vary widely. Many people in all cultures are still intimidated by computer technology. Therefore, it will be important to design easy-to-use and user-friendly user interfaces for any assessment software that might be used. The design(s) will need to be non-culturally specific, and will need to be accessible to a wide range of people with disabilities and to people who are not literate.

But hardware, compression software, and well designed graphical user interfaces, while necessary, are not sufficient. In order to be successful there also needs to be software that recognizes speech and writing in the target language(s) and turns it into text, as well as the development of performance analysis software for both speaking and writing. Such software, while still somewhat experimental, is now commercially available,

especially for writing. It is a mixture of artificial intelligence, statistical modeling, natural language processing, and proven assessment techniques for language performance skills. (See Fig. 1).

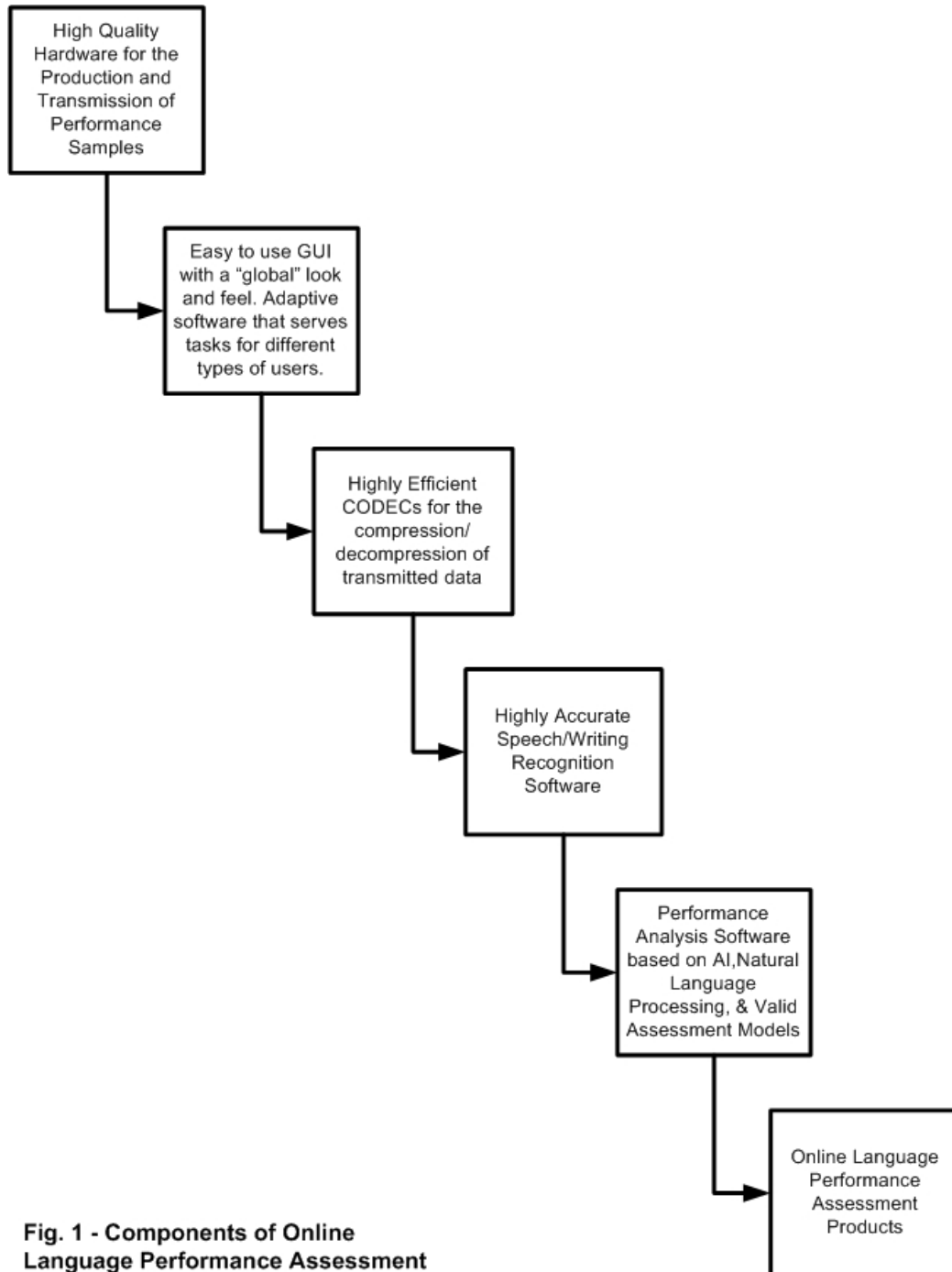


Fig. 1 - Components of Online Language Performance Assessment

Technologies for the Online Assessment of Speaking

In addition to the technological hurdles that have had to be overcome, the online assessment of speaking remains difficult because speaker responses cannot be easily predicted. At the present time, it is only possible to assess online and give immediate feedback on some of the sub-skills of speaking.

Currently, the most common way of using online technology for speaking assessment is to collect speech samples, and have them transmitted via the Internet for evaluation by trained graders or by peers. An online public speaking course at Columbus State University (Rosenthal, 2000) used "virtual adaptive learning technology" text based on Howard Gardner's Theory of Multiple Intelligences and streaming video in speech assignments that were evaluated by both the instructor and peers. Davis (2003) reports that the Center for Applied Second Language Studies (CASLS) at the University of Oregon is developing an integrated online assessment of speaking, listening, reading and writing using streaming media. Brigham Young University has a computerized program for that assessment of oral skills that allows the teacher "to review instantaneously all or part of any student's responses in any order desired" (Larson, 1998).

The next step is to have computerized feedback on speech performance. A consortium of the American Council on the Teaching of Foreign Languages, Brigham Young University, the Center for Applied Linguistics, the Defense Language Institute, and San Diego State University are collaborating in an effort to develop an oral screening test. "This test would allow foreign language and area studies graduate program applicants, pre-service teachers, candidates to programs in translation & interpretation, or self-instructional learners an opportunity to take an online oral interview consisting of automated audio and visual elicitation prompts that could be used to provide informative feedback regarding their oral ability" (LARC-SDSU, 2004). Frase (2002)

reports that Ordinate's "PhonePass Spoken English Test" is the first automatic test *by telephone* for people who speak English as a foreign language. Frase adds, "Given the current state of speech technology and the features that such technologies can recognize, it is clear that they can be used to replace a human judge in tasks that yield indirect measures of communicative skill. Tasks that constrain the domain of responses can be created to make assessments quite accurate. For instance, indirect measures of communicative skill can be derived from performance on tasks like sentence repetition or sentence completion."

It is difficult to assess speaking online without good speech recognition software. Currently, software such as Dragon Systems Naturally Speaking, and IBM's ViaVoice can recognize human speech at a 95-97% accuracy rate if the speaker spends at least a half hour reading a speech sample for the software. Speech recognition software is based on statistical Markov models that continue to be refined. Once speech has been turned into text, it can be analyzed for content, spelling, and grammar using other statistical techniques such as Latent Semantic Analysis (LSA).

Latent Semantic Analysis is a theory and mathematical model for analyzing text for context and meaning (Landauer and Dumais, 1997). "The underlying idea is that the aggregate of all the word contexts in which a given word does and does not appear provides a set of mutual constraints that largely determines the similarity of meaning of words and sets of words to each other." LSA is also one of the main methods used in computerized analysis of writing.

Current Technologies for the Online Assessment of Writing

The online assessment of writing is also difficult, although it seems to be easier than the online assessment of speaking. Perhaps this is because there has been more work done on the analysis of writing, making it easier to develop algorithms for assessment. For example, the Computerized Inventory of Developmental Writing Traits (CIDWT) has been available for over a dozen years (McCurry, 1992, cited in US Dept of Education, 2000). According to the US Department of Education (2000), the CIDWT has been found "to correlate highly with teacher ratings" of writing. At the same time, the authors of this report contend that "rating of qualities such as organization, tone of voice, originality of ideas, etc. are not readily conducive to computerized scoring." Yet progress is being made. For example, "centering theory" is being used to detect "rough shifts" in student essays, where the student abruptly changes from one topic to another (Miltakaki and Kukich, 2000). MacKay and Emerson (2000) describe the online testing of grammar, spelling and punctuation using multiple choice questions. Pickard (1998) shows how a concordancer can be used to analyze journals and course logs.

Automated essay scoring has a long history, starting with work by Ellis Page in the 1960s. His program Project Essay Grade (PEG) has shown that it can achieve reliability correlations comparable with human judges. (Rudner and Gagne, 2001). This is how it works:

"The underlying theory is that there are intrinsic qualities to a person's writing style called trins that need to be measured, analogous to true scores in measurement theory. PEG uses approximations of these variables, called proxes, to measure these underlying traits. Specific attributes of writing style, such as average word length, number of semicolons, and word rarity are examples of proxes that can be measured directly by PEG to generate a grade. For a given sample of essays, human raters grade a large number of essays (100 to 400), and determine values for up to 30 proxes. The grades are then entered as the criterion variable in a regression equation with all of the proxes as predictors, and beta weights are computed for each predictor.

For the remaining unscored essays, the values of the proxies are found, and those values are then weighted by the betas from the initial analysis to calculate a score for the essay." (Rudner and Gagne, 2001)

A more recent statistical technique for essay scoring, called Latent Semantic Analysis, has gained a strong following, and is the basis for a number of commercial products. It was originally designed for indexing documents for information retrieval. LSA uses "calibration documents" as models and looks for similarities with the essay being evaluated. Foltz, Kintsch and Landauer (1998) report very high correlations between LSA scores and human rated essays. According to Lee and Foltz (1998), LSA is being applied to intelligent tutoring systems that teach writing.

A third technology for judging writing is the *e-rater* program developed by the Educational Testing Service (ETS). It combines features of both PEG and LSA, and also looks at "syntactic variety". (Burstein, 2003)

Finally, a technology that is worth examining is computerized *adaptive* testing:

Computerized *adaptive* testing (CAT) is a unique type of CBT [computer-based testing] which enables greater precision and efficiency in some forms of assessment by first estimating an examinee's proficiency level (typically on the basis of initial item responses) and then adapting to it, presenting only those items that are expected to give the most information about that individual (i.e., neither overly easy nor overly difficult items). It has also been argued that CATs will prove advantageous for certain language assessment purpose. (Norris, 2001)

These are the principal technologies available today for the online assessment of speaking and writing. In the near future, artificial intelligence and continuous measurement techniques will extend these technologies into highly accurate assessment tools. We are just beginning to see this trend with the announcement of new commercial products for the online assessment of speaking and writing.

Commercial Tools for the Assessment of Speaking and/or Writing

The technologies described above are starting to come out of the laboratory and be commercialized. Here is a brief alphabetical listing of commercial products that provide computerized online assessment of speaking and/or writing, along with a brief description from each company/organization's publicity materials:

Criterion – Educational Testing Service – based on *e-rater* technology

Web site: <http://criterion.ets.org/cwe/student/>

“Criterion is a web-based service that evaluates a student's writing skills and provides instant score reporting and diagnostic feedback to both the instructor and student. The new release offers significant enhancements in functionality and design.

Criterion Online Writing Evaluation utilizes automated scoring technologies to provide writers with instant score and writing analysis feedback on writing samples they submit. Licensed to academic institutions and related service networks, Criterion offers writers an opportunity to improve skills by writing and revising in a self-paced, risk-free environment.” (Business Wire, 2002)

DIALANG – developed with the support of the European Commission

Web site: www.dialang.org

“DIALANG offers carefully designed and validated tests of different language skills, together with a range of feedback and expert advice on how to improve your skills.

DIALANG also offers scientifically validated self-assessment activities, because it is now widely recognized that being able to judge your own language proficiency is an important part of learning a language.”

According to Noijons (1999), clients choose to be assessed on reading, listening, speaking, structures/grammar, or vocabulary.

IntelliMetric – from Vantage Learning

Web Site: www.vantagelearning.com

IntelliMetric(TM), Enterprise Edition, is a faster version of Vantage Learning’s award-winning IntelliMetric essay scoring technology, for use in high-volume assessment programs. Designed for programs needing to score more than 200,000 essay responses annually, IntelliMetric Enterprise Edition scores examinee responses to essay questions faster and more accurately than all competitive products on the market. Recently, the IntelliMetric system scored 175,000 benchmark tests in 48 hours. (Business Wire, 2004a, 2004b).

Two other products, ACCUPLACER’s *WritePlacer*, and McGraw-Hill’s *Writing RoadMap* uses the same IntelliMetric engine.

Language Evaluation and Assessment Project (LEAP) – University of Toronto

Web site: www.lang.utoronto.ca/LEAP/

LEAP consists of the following tests:

- A. Grammar: multichoice questions (parts A1 and A2)
- B. Oral discrimination: multichoice questions
- C. Reading comprehension: multichoice questions
- D. Writing: 135-150 words essay
- E. Oral comprehension: multi-choice questions and 'dictée' (parts E1, E2 and E3)

LearnDirect – UK company with language diagnostics including writing

Web site: www.learndirect.org

LearnDirect offers language instruction, with a computerized diagnostic assessment to ascertain ones level. The diagnostic assessment includes writing. The program can be downloaded from the LearnDirect Web site.

Odyssey Writer - from CompassLearning

Web site: <http://www.compasslearningodyssey.com/>

“The product breaks down the writing process into simple, more accessible steps like prewriting, drafting, revising, and publishing.” It is flexible - students can use any or all of its components, in the order and method that best suits their own work styles.

Odyssey Writer features assessment tools that encompass analytic and holistic rubrics for use by students and teachers. These tools (note card outliner, graphic organizer, drafting instrument, and assessment rubrics) not only provide the support for students to learn, they help teachers actively facilitate and assess each student's writing. Through a direct online prompt, students are given an assignment, work on it, and then self-assesses their own work. Teachers can meet each student's individual needs by and through the use of a variety of structural tools. For instance, teachers can give feedback as well as monitor students' self-assessments; multiple projects can be worked on simultaneously; and students are able return to projects, from school or home, and continue to work on them. Digital portfolios allow the teacher, student or parent to access these portfolios. The Self-Assessment Tool encourages students to continually revise their writing, carrying out self-evaluation of their work using a predetermined rubric. Students can compare their work side-by-side to the rubric criteria, adding, substituting, deleting or moving sentences and paragraphs to produce "quality" work.” - PR Newswire (2004)

Speech Tools 2.0 – from SIL Software (SIL is a service organization that works with people who speak the world’s lesser-known languages. Based in Dallas, TX)

Web Site: <http://www.sil.org/computing/speechtools/>

Components of Speech Tools include:

- Speech Analyzer
- Speech Manager
- IPA Help

Demo versions of Speech Tools can be downloaded for evaluation.

The purpose of this working paper was to identify technologies, products and researchers in the field of online assessment of speaking and writing. The next step is to undertake a detailed evaluation of the findings, and to use the bibliography to identify further resources.

* * *

Dr. Gary Woodill is Chief Learning Officer, Operitel Corporation. A former teacher and university professor in education, he is also a writer and speaker on e-learning and knowledge management. For the past 12 years he has been involved in the design of learning management systems and online courses. Other white papers by Dr. Woodill can be found at www.learnflex.com. You can reach him at gwoodill@operitel.com.

References and Bibliography

Atkinson, T. and Davies, G. (2004) *Computer Aided Assessment (CAA) and Language Learning*. Information and Communications Technology for Language Teachers (ICT4LT) Course Module 4.1: Information and Communications Technology for Language Teachers. Available at: http://www.ict4lt.org/en/en_mod4-1.htm#2.3.

Bernstein, J. and Hirschman, L. (2001) *Evolution of Performance Measures for Language Technologies*. Paper presented at the 23rd Annual Language Testing Research Colloquium, St. Louis, Missouri. Feb. 20-24, 2001.

Brown, A., Iwashita, N., McNamara, T. and O'Hagan, S. (2001) *Investigating Raters' Orientation in Specific-purpose Task-based Oral Assessment*. Paper presented at the 23rd Annual Language Testing Research Colloquium, St. Louis, Missouri. Feb. 20-24, 2001.

Brown J. (1997) "Computers in language testing: present research and some future predictions", *Language Learning and Technology* 1, 1: 44-59. Available at: <http://llt.msu.edu/vol1num1/brown/>

Burstein, Jill. (2003). The *E-rater*® Scoring Engine: Automated Essay Scoring With Natural Language Processing. In *Automated essay scoring: A cross-disciplinary perspective*. Edited by Mark D. Shermis and Jill Burstein, Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Burstein, Jill, Marcu, Daniel, and Knight, Kevin (2003). Finding the WRITE Stuff: Automatic Identification of Discourse Structure in Student Essays. In S. Harabagiu and F. Ciravegna (Eds.) *Special Issue on Advances in Natural Language Processing, IEEE Intelligent Systems*, v.18, n.3, pp. 32-39: www.computer.org/intelligent/archives.htm.

Burstein, Jill C. and Chodorow, Martin. (2002). Directions in Automated Essay Analysis, In *The Oxford Handbook of Applied Linguistics*. Edited by Robert Kaplan, Oxford University Press, New York.

Burstein, J., Chodorow, M., & Leacock, C. (2003). CriterionSM: Online essay evaluation: An application for automated evaluation of student essays. *Proceedings of the Fifteenth Annual Conference on Innovative Applications of Artificial Intelligence*, Acapulco, Mexico, August 2003. Available at:
<http://www.ets.org/research/dload/iaai03bursteinj.pdf>.

Burstein, Jill. and Marcu, Daniel. (2003). Developing Technology for Automated Evaluation of Discourse Structure in Student Essays. In *Automated essay scoring: A cross-disciplinary perspective*. Edited by Mark D. Shermis and Jill Burstein, Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1989). The four generations of computerized educational measurement. In R. L. Linn (Ed.), *Educational Measurement: Third edition*. (pp. 367-407) New York, NY: Macmillan Publishing Company.

Business Wire. (2000) ETS Technologies, Inc. and PLATO Learning, Inc. To Offer Online Writing Assessment. *Business Wire*. New York: Nov 15, 2000.

Business Wire. (2001) LearnATest.com Now Individually Scores Online Writing Assessments. *Business Wire*, New York: Jan 12, 2001.

Business Wire. (2002a) ETS Technologies Introduces the Next Generation of Online Writing Assessment Tools; Criterion Provides Students Online Writing Practice and Immediate Feedback. *Business Wire*, New York: Sep 18, 2002.

Business Wire. (2002b) CTB/McGraw-Hill Launches Writing Roadmap Online Writing Assessment; New Tool Helps to Enhance Students' Written Communication Skills. *Business Wire*. New York: Aug 29, 2002.

Business Wire. (2004a) Vantage Learning Scores and Reports 175,000 Student Benchmark Tests in Less than 48 Hours. *Business Wire*. New York: Feb 16, 2004.

Business Wire. (2004b) Vantage Learning Announces Release of Fastest Automated Essay Scoring Technology: IntelliMetric Enterprise Edition. *Business Wire*, New York: Feb 9, 2004.

Carr, Nathan. (2001) *Construct Validation of an Integrated Communicative Language Test*. Paper presented at the 23rd Annual Language Testing Research Colloquium, St. Louis, Missouri. Feb. 20-24, 2001.

Conference on College Composition and Communication (CCCC) (2004) Writing assessment: a position statement. *College Composition and Communication Online*. Online newsletter- <http://www.ncte.org/about/over/positions/category/write/107610.htm>

Cebrian, B. and Grellett, P. (1998) Using and integrating CALL and multiple media for specific purposes in the teaching and learning process. Paper presented at the Inaugural WorldCALL Conference, Stream Five: Towards new models of assessment and evaluation in CALL. Abstract found at: www.hlc.unimelb.edu.au/worldcall/str5.html.

Center for Education (CFE) (2002) *Technology and Assessment: Thinking Ahead -- Proceedings from a Workshop*. Board on Testing and Assessment, National Research Council.

Chalhoub-Deville, Micheline. (Ed.) (1999) *Issues in Computer-Adaptive Testing of Reading Proficiency*. Cambridge: University of Cambridge.

Chalhoub-Deville, Micheline. (2001) Language testing and technology: past and future. *Language Learning and Technology*, v. 5, n. 2, May.

Cho, Yeonsuk. (2001) *Examining a Process-oriented Writing Assessment for Large Scale Assessment*. Paper presented at the 23rd Annual Language Testing Research Colloquium, St. Louis, Missouri. Feb. 20-24, 2001.

Chodorow, Martin, Leacock, Claudia and Miller, George. (2000) "A topical/local classifier for word sense identification". *Computers and the Humanities* 34:1-2.

Clark, Damien. (2004) Enhancing the IMS Q&TI Specification by adding support for dynamically generated parameterised quizzes. Thesis Abstract. Central Queensland University. Available at:

http://infocom.cqu.edu.au/Staff/Damien_Clark/Research/thesis.html

Cooper, T. and Love, T. (2004) *Online portfolios: issues of assessment and pedagogy*. Web article. Edith Cowan University, Perth, Western Australia. Available at

<http://www.aare.edu.au/01pap/coo01346.htm>.

Davis, Robert (2003) *Online proficiency testing-SPANISH*. e-Mail communication. University of Oregon.

Dunkel P. (1999) "Considerations in developing or using second / foreign language proficiency computer-adaptive tests", *Language Learning & Technology* 2, 2: 77-93.

Available at: <http://ilt.msu.edu/vol2num2/article4/>

Eschenbach, Elizabeth. (2001) *Improving Technical Writing via Web-based Peer Review of Final Reports*. Conference Presentation, ASEE/IEEE Frontiers in Education Conference.

Flexible Education Unit, University of Tasmania. (2004) *Assessment and online discussion: Eight ways to incorporate online discussion into assessment*. Online newsletter, March 9, 2004. Available at:
http://www.utas.edu.au/teachingonline/develop/webct_tools/communications/assess_online_discussion.doc.

Foltz, P. W., Laham, D. & Landauer, T. K. (1999). Automated Essay Scoring: Applications to Educational Technology. In *Proceedings, EdMedia '99*. Available at:
<http://www-psych.nmsu.edu/~pfoltz/reprints/Edmedia99.html>.

Frase, Lawrence. (2002) Seven technologies for assessment. Online paper at:
<http://mason.gmu.edu/~amelmed/768%20Fall%2002/Frase.htm>.

Goodwin-Jones, Robert. (2001) Emerging technologies: language testing tools and technologies. *Language Learning and Technology*, v. 5, n. 2, May. Available at:
<http://llt.msu.edu/vol5num2/emerging/default.html>.

Gupta, P., Schulze, M. and Davies, G. (2004) *Human Language Technologies*. Information and Communications Technology for Language Teachers (ICT4LT) Course Module 4.1: Information and Communications Technology for Language Teachers. Available at: http://www.ict4lt.org/en/en_mod3-5.htm

Hawkey, R., Saville, N. and Banks, C. (2001) *Towards a Common Performance Description Scale for Non-native Speaker Writing*. Paper presented at the 23rd Annual Language Testing Research Colloquium, St. Louis, Missouri. Feb. 20-24, 2001.

Hill, K. and Gruba, P. (1998) Technologically speaking: Delivery alternatives for the assessment of speaking. Paper presented at the Inaugural WorldCALL Conference, Stream Five: Towards new models of assessment and evaluation in CALL. Abstract found at: www.hlc.unimelb.edu.au/worldcall/str5.html.

Huston, Michelle. (2000) *Online Course Materials: Asian language learning*. Paper presented at AusWeb01, the Seventh Australian World Wide Web Conference, 21st-25th April, 2000. Available at: <http://ausweb.scu.edu.au/aw01/papers/refereed/huston/paper.html>.

Kenyon, D. and Malabonga, V. (2001) Comparing examinee attitudes toward computer-assisted and other oral proficiency assessments. *Language Learning and Technology*, v. 5, n. 2, May. Available at: <http://llt.msu.edu/vol5num2/kenyon/default.html>.

Knoy, T., Lin, S.J., Liu, Z.F., and Yuan, S.M. (2001) *Networked Peer Assessment in Writing: Copyediting skills instruction in an ESL Technical Writing Course*. Paper presented at the International Conference on Computers in Education, 2001. Available online at: http://www.cc.nctu.edu.tw/~tedknoy/html/News_2001_NPAW.htm.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes*, **25**, 259-284.

LARC-SDSU - Language Acquisition Resources Center at San Diego State University. (2004) Projects: CAST (Computerized ACTFL-based Screening Tool). Web page at <http://larcnet.sdsu.edu/testing.php?page=cast>

LANG – Language Advisory Networking Group, University of Toronto (2004) *LEAP: Language Evaluation and Assessment Project*. Case Study in French available at:
http://www.lang.utoronto.ca/pt_prototype/

Larkey, L.S. (1998) *Automated essay grading using text categorization techniques*. Paper presented at the 21st International Conference of the Association for Computing Machinery-Special Interest Group on Information Retrieval, Melbourne, Australia.

Larson, Jerry. (1998) *Using the computer for oral skills assessment*. Paper presented at the Inaugural WorldCALL Conference, Stream Five: Towards new models of assessment and evaluation in CALL. (Presented at the conference by Kim Smith). Abstract found at: www.hlc.unimelb.edu.au/worldcall/str5.html.

Lee, A. and Foltz, P. (1998) *KDI: Learning through writing using adaptive tutoring systems: modeling knowledge representations from open-ended questions*. National Science Foundation Award Abstract. Available at:
<https://www.fastlane.nsf.gov/servlet/showaward?award=9873491>.

LeLoup, J. and Ponterio, R. (2001) On the Net: language testing resources. *Language Learning and Technology*, v. 5, n. 2, May. Available at:
<http://llt.msu.edu/vol5num2/onthenet/default.html>.

Luoma, Sari. (2003) *Assessing Speaking*. Cambridge, England: Cambridge University Press.

MacKay, E. and Emerson, L. (2000) *Web-based assessment of writing skills*. Online article, Massey University, New Zealand. Available at:
<http://www.lboro.ac.uk/service/fli/flicaa/conf2000/pdfs/mackayb.pdf>

Miceli, T. and Kennedy, C. (1998) Student talk and multimedia. Paper presented at the Inaugural WorldCALL Conference, Stream Five: Towards new models of assessment and evaluation in CALL. Abstract found at:

www.hlc.unimelb.edu.au/worldcall/str5.html.

Miller, George and Leacock, Claudia. (2000) Lexical representations for sentence processing. In Y. Ravin and C. Leacock, editors, *Polysemy: Theoretical and Computational Approaches*. Oxford University Press, Oxford.

Miller, Tristan. (2003a) Essay Assessment with Latent Semantic Analysis. *Journal of Educational Computing Research*, 28. Available online at: <http://www.dfki.uni-kl.de/~miller/publications/miller03a.pdf>.

Miller, Tristan. (2003b) *Generating Coherent Extracts of Single Documents using Latent Semantic Analysis*. Masters Thesis. University of Toronto. Available online at: <ftp://ftp.cs.toronto.edu/pub/gh/Miller-thesis.pdf>.

Miltsakaki, E. and Kukich, K. (2000) Automated Evaluation of Coherence in Student Essays. In *Proceedings of the Workshop on Language Resources and Tools in Educational Applications*, LREC 2000, pages 7-14, Athens. Available at: <http://www.ling.upenn.edu/~elenimi/lrecpaper.pdf>.

Mislevy, Robert J. and Almond, Russell G. (2002). "Design and analysis in task-based language assessment." *Language Testing* 19.4: 477-496

Noijons, José. (1999) DIALANG: diagnostic language testing through the internet. Paper presented at EuroCALL 99. Abstract available at: <http://www.languages.dk/eurocall/eurocall99/recordings/dialang.htm>.

Norris, John. (2001) Concerns with computerized adaptive oral proficiency assessment. *Language Learning and Technology*, v. 5, n. 2, May. Available at: <http://llt.msu.edu/vol5num2/norris/default.html>.

Pickard, Valerie. (1998) Concordancing student journal entries: An aid to course evaluation. Paper presented at the Inaugural WorldCALL Conference, Stream Five: Towards new models of assessment and evaluation in CALL. Abstract found at: www.hlc.unimelb.edu.au/worldcall/str5.html.

Powers, Donald E., Jill C. Burstein, Martin Chodorow, Mary E. Fowles, and Karen Kukich (2002). Stumping *e-rater*: challenging the validity of automated essay scoring. *Computers and Human Behavior*, 18, pp. 103-13.

PR Newswire (2004) Odyssey Writer Gives Students the Tools and Skills to Become Strong Writers. *PR Newswire*, NY Mar 17, 2004.

Ram, A. and Moorman, K. (Eds.) (1999) *Understanding Language Understanding: computational models of reading*. Cambridge, Massachusetts: MIT Press.

Rendall, H. and Davies, G. (2004) *Using Text Tools in the Modern Foreign Languages Classroom*. Information and Communications Technology for Language Teachers (ICT4LT) Course Module 4.1: Information and Communications Technology for Language Teachers. Available at: http://www.ict4lt.org/en/en_mod1-3.htm.

Roever, Carsten. (2001) Web-based language training. *Language Learning and Technology*, v. 5, n. 2, May. Available at: <http://llt.msu.edu/vol5num2/roever/default.html>.

Rosenthal, Ann. (2000) Accountability and Assessment in an Online Public Speaking Course: Is That Possible? *PROCEEDINGS: Advocacy, Accountability and Assessment*. V69, Spring 2000, URL: <http://www.valdosta.edu/comarts/Rosenthal.html#Rosenthal00>

Rudner, L. and Gagne, P. (2001) An overview of three approaches to scoring written essays by computer. *Practical Assessment, Research and Evaluation*, v. 7, n. 26. Article from an electronic journal found at: <http://pareonline.net/getvn.asp?v=7&n=26>.

Sawaki, Yasuyo. (2001) Comparability of conventional and computerized tests of reading in a second language. *Language Learning and Technology*, v. 5, n. 2, May. Available at: <http://ilt.msu.edu/vol5num2/sawaki/default.html>.

Shermis, M. D., & Burstein, J. (Eds.) (2003). Automated essay scoring: A cross-disciplinary perspective. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Stauffer, S and Kenyon, D. (2001) *A Computer-assisted, Computer-adaptive Oral Proficiency Assessment Instrument Prototype*. Paper presented at the 23rd Annual Language Testing Research Colloquium, St. Louis, Missouri. Feb. 20-24, 2001.

Strong-Krause, Diane. (2001) *A Tree-based Modeling Approach to Construct Validation of a Computer-delivered Speaking Test*. Paper presented at the 23rd Annual Language Testing Research Colloquium, St. Louis, Missouri. Feb. 20-24, 2001.

Syllabus IT Trends. (2004) University of Houston: Online Writing Assessment enables Effective, Efficient Testing. *Syllabus IT Trends Newsletter (Online)*, Mar. 25, 2004.

Tan, Sophia Huey Shan. (1998) In tune with technology and testing. Paper presented at the Inaugural WorldCALL Conference, Stream Five: Towards new models of assessment and evaluation in CALL. Abstract found at: www.hlc.unimelb.edu.au/worldcall/str5.html.

Taylor, L. and Beeston, S. (2001) *Revising Instruments for Rating Speaking: Combining Qualitative and Quantitative Insights*. Paper presented at the 23rd Annual Language Testing Research Colloquium, St. Louis, Missouri. Feb. 20-24, 2001.

United States Department of Education, National Center for Education Statistics. (2000) *The NPEC Sourcebook on Assessment, Volume 1: Definitions and Assessment Methods for Critical Thinking, Problem Solving, and Writing*, Washington, DC: U.S. Government Printing Office.

Weiner, Howard. (Ed.) (2000) *Computerized Adaptive Testing: a primer*. Mahwah, NJ: Lawrence Erlbaum.

Weir, C., O'Sullivan, B. and French, A. (2001) *Task Difficulty in Testing Spoken Language: A Socio-cognitive Perspective*. Paper presented at the 23rd Annual Language Testing Research Colloquium, St. Louis, Missouri. Feb. 20-24, 2001.

APPENDIX: Web-Sites Related to Computer –based Language Assessment

<http://carla.acad.umn.edu/CAT.html> (University of Minnesota, Center for Advanced Research on Language Acquisition; computer-adaptive language testing project)

<http://ericae.net/scripts/cat/catdemo.htm> (ERIC Clearinghouse on Assessment and Evaluation; interactive CAT tutorial)

<http://web.uvic.ca/hrd/halfbaked/#latest> (Hot Potatoes; computer-based testing software and information)

<http://www.ets.org/cbt/index.html> (Educational Testing Service; computer-based testing information)

<http://www.lll.hawaii.edu/nflrc/cbt.html> (University of Hawaii, National Foreign Language Resource Center; research and development on computer-based tests for less commonly taught languages)

<http://www.rasch.org/index.htm> (Institute for Objective Measurement; software and information on Item Response Theory)

<http://www.toefl.org/cbtindex.html> (computer-based Test of English as a Foreign Language; main information page and example test items)

<http://www2.hawaii.edu/~roever/wbt.htm> (information and links on Web-based language testing)

<http://www.dundee.ac.uk/languagestudies/ltest/ltr.html> (Resources in Language Testing)