

A BIBLIOGRAPHY OF MACHINE SCORING OF STUDENT WRITING, 1962–2005

Richard H. Haswell

This bibliography focuses on the theory, design, application, and implications of automated or machine rating of extended student writing. It covers the literature in English that speculates and reports on the success of computers in scoring “free text” or essaylike compositions of students. The writing scored may range from paragraphs answering examination questions to drafts for course papers to timed extemporaneous essays written under formal assessment conditions.

As such, the bibliography treats only incidentally other areas of machine analysis of writing: application of readability formulas, text mining, information retrieval, computer-assisted instructional feedback systems, programmed learning systems, automatic text generation, machine translation, computerized record keeping, and spelling-, grammar-, and style-checkers. The bibliography also eschews conference papers that are not accessible through libraries or the Internet, journalistic items such as press releases and news stories, and product ads and plugs.

The terminus ad quo is 1962, with Walter R. Reitman’s thoughts on the possibility of machine grading at a conference on needed research in the teaching of English. The terminus ad quem is 2005, although the representation of work from that year is thin. None of the contributions to this present volume are included here.

- Ajay, Helen Beuck. 1973. Strategies for Content Analysis of Essays by Computer. Ph.D. diss., University of Connecticut, Storrs.
- Ajay, Helen Beuck, and Ellis Batten Page. 1973. *Analysis of Essays by Computer (AEC-II)*. Final report to the National Center for Educational Research and Development. Washington, DC: Office of Education, Bureau of Research.
- Anderson, C. W., and G. E. McMaster. 1982. Computer Assisted Modeling of Affective Tone in Written Documents. *Computers and the Humanities* 16:1–9.
- Andeweg, Bas A., J. C. de Jong, and R. Natadarma. 1996. Improving Writing Skills through Ganesh. In *Effective Teaching and Learning of Writing: Current Trends in Research*, edited by Gert Rijlaarsdam, Huub van den Bergh, and Michel Couzijn. Amsterdam: Amsterdam University Press.

- Anson, Chris M. 2002. Responding to and Assessing Student Writing: The Uses and Limits of Technology. In *Teaching Writing with Computers: An Introduction*, edited by Pamela Takayoshi and Brian Huot. Boston: Houghton Mifflin.
- Attali, Yigal. 2004. *Exploring the Feedback and Revision Features of Criterion*. Princeton, NJ: Educational Testing Service. www.ets.org/research/dload/NCME_2004Attali.pdf.
- Baron, Dennis. 1998. When Professors Get A's and the Machines Get F's. *Chronicle of Higher Education*, November 20, A56.
- Barth, Rodney J. 1979. ERIC/RCS Report: An Annotated Bibliography of Readings for the Computer Novice and the English Teacher. *English Journal* 68 (1): 88–92.
- Bejar, Isaac I. 1996. *Generative Response Modeling: Leveraging the Computer as a Test Delivery Medium*. English Report No. ETS-RR-96-13. Princeton, NJ: Educational Testing Service. ERIC Document Reproduction Service, ED 401 302.
- Bennett, Randy Elliott 1993. Toward Intelligent Assessment: An Integration of Constructed-Response Testing, Artificial Intelligence, and Model-Based Measurement. In *Test Theory for a New Generation of Tests*, edited by Norman Frederiksen, Robert J. Mislevy, and Isaac I. Bejar. Hillsdale, NJ: Lawrence Erlbaum.
- Bennett, Randy E., and Isaac I. Bejar. 1998. Validity and Automated Scoring: It's Not Only the Scoring. *Educational Measurement: Issues and Practice* 14 (4): 9–12.
- Bereiter, Carl. 2003. Foreword to Shermis and Burstein 2003, vii x.
- Bergler, Sabine. 1995. From Lexical Semantics to Text Analysis. In *Computational Lexical Semantics*, edited by Patrick Saint-Dizier and Evelyne Viegas. Cambridge: Cambridge University Press.
- Breland, Hunter M. 1996. Computer-Assisted Writing Assessment: The Politics of Science versus the Humanities. In *Assessment of Writing: Politics, Policies, Practices*, edited by Edward M. White, William D. Lutz, and Sandra Kamusikiri. New York: MLA.
- Breland, Hunter M., and Eldon G. Lytle. 1990. Computer-Assisted Writing Skill Assessment Using WordMAP. Paper presented at the American Educational Research Association, Boston, April. www.writinglab.com/brelandlytle.html.
- Brock, Mark N. 1995. Computerised Text Analysis: Roots and Research. *Computer Assisted Language Learning* 8 (2–3): 227–258.
- Burstein, Jill. 1999. Enriching Automated Essay Scoring Using Discourse Marking. In *COLING-ACL '98: Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Aug. 10–14, 1998*. Montreal: Association for Computational Linguistics.
- . 2003. The E-rater Scoring Engine: Automated Essay Scoring with Natural Language Processing. In Shermis and Burstein 2003.
- Burstein, Jill, and Martin Chodorow. 1999. Automated Essay Scoring for Nonnative English Speakers. In *37th Annual Meeting of the Association for Computational Linguistics: Proceedings of the Conference: 20–26 June 1999, University of Maryland, College Park, Maryland, USA*. New Brunswick, NJ: Association for Computational Linguistics. www.ets.org/research/dload/ac199rev.pdf.
- Burstein, Jill, Lawrence T. Frase, April Ginther, and Leslie Grant. 1996. New Technologies for Language Assessment. In *Annual Review of Applied Linguistics*, vol. 16, edited by Robert B. Kaplan, Charles A. Ferguson, Henry G. Widdowson, Richard Tucker, and Merrill Swain. Cambridge: Cambridge University Press.
- Burstein, Jill, and Derrick Higgins. 2005. Advanced Capabilities for Evaluating Student Writing: Detecting Off-Topic Essays without Topic-Specific Training. Paper presented at the International Conference on Artificial Intelligence in Education, July 2005, Amsterdam. www.ets.org/Media/Research/pdf/erater_burstein_higgins_CR.pdf.
- Burstein, Jill, and Randy M. Kaplan. 1995. On the Application of Context to Natural Language Processing Applied to the Analysis of Test Responses. In *IJCAI: Proceedings from the Workshop on Context in Natural Language Processing*. Montreal: International Joint Conference on Artificial Intelligence.

- Burstein, Jill, Karen Kukich, L. Braden-Harder, Martin Chodorow, S. Hua, and Bruce Kaplan. 1998. *Computer Analysis of Essay Content for Automatic Score Prediction: A Prototype Automated Scoring System for GMAT Analytical Writing Assessment*. Research Report 98-15. Princeton, NJ: Educational Testing Service.
- Burstein, Jill, Karen Kukich, Susanne Wolff, Chi Lu, and Martin Chodorow. 1998. *Computer Analysis of Essays*. Princeton, NJ: Educational Testing Service.
- . 1998. Enriching Automated Scoring Using Discourse Marking. Paper presented at the workshop on discourse relations and discourse marking at the annual meeting of the Association for Computational Linguistics, Montreal. www.ets.org/research/erater.html.
- Burstein, Jill, Karen Kukich, Susanne Wolff, Chi Lu, Martin Chodorow, L. Braden-Harder, and M. D. Harris. 1998. Automated Scoring Using a Hybrid Feature Identification Technique. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. Montreal: Association for Computational Linguistics. www.ets.org/research/erater.html.
- Burstein, Jill, and Daniel Marcu. 2000. Benefits of Modularity in an Automated Essay Scoring System. Paper presented at the Workshop on Using Toolsets and Architectures to Build NLP Systems, 18th International Conference on Computational Linguistics, Luxembourg, August 2000. ftp.ets.org/pub/res/erater_colinga4.pdf.
- . 2000. Toward Using Text Summarization for Essay-Based Feedback. Paper presented at the Conference TALN 2000, Lausanne. www.isi.edu/~marcu/papers/talnsum2000.pdf.
- . 2003. Automated Evaluation of Discourse Structure in Student Essays. In Shermis and Burstein 2003.
- . 2004. A Machine Learning Approach for Identification of Thesis and Conclusion Statements in Student Essays. *Computers and the Humanities* 37 (4): 455–467.
- Burstein, Jill, Daniel Marcu, S. Andreyev, and Martin Chodorow. 2001. Towards Automatic Classification of Discourse Elements in Essays. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*. Montreal: Association for Computational Linguistics.
- Burstein, Jill, Susanne Wolff, and Chi Lu. 1999. Using Lexical Semantic Techniques to Classify Free-Responses. In *Breadth and Depth of Semantic Lexicons*, edited by Evelyne Viegas, 227–246. Dordrecht and Boston: Kluwer Academic Press.
- Byerly, Gayle A. 1978. CAI in College English. *Computers and the Humanities* 12 (3): 281–285.
- Calfee, Robert. 2000. To Grade or Not to Grade. *IEEE Intelligent Systems* 15 (5): 35–37. www.knowledge-technologies.com/papers/IEEEdebate.pdf.
- Chase, Clinton I. 1999. *Contemporary Assessment for Educators*. New York: Addison-Wesley Longman.
- Cheville, Julie. 2004. Automated Scoring Technologies and the Rising Influence of Error. *English Journal* 93 (4): 47–52.
- Chodorow, Martin, and Claudia Leacock. 2000. An Unsupervised Method for Detecting Grammatical Errors. In *Proceedings of the First Conference of the North American Chapter of the Association for Computational Linguistics*. ACM International Conference Proceedings Series, Vol. 4. New Brunswick, NJ: Association of Computational Linguistics.
- Chung, Gregory K. W. K., and Eva L. Baker. 2003. Issues in the Reliability and Validity of Automated Scoring of Constructed Responses. In Shermis and Burstein 2003.
- Chung, Gregory K. W. K., and Harold F. O'Neil, Jr. 1997. *Methodological Approaches to Online Scoring of Essays*. Report No. CSE-TR-461. Los Angeles: University of California, Center for the Study of Evaluation, Standards, and Student Testing. ERIC Document Reproduction Service, ED 418 101.
- Cizek, Gregoy J., and Bethany A. Page. 2003. The Concept of Reliability in the Context of Automated Essay Scoring. In Shermis and Burstein 2003.
- Clauser, Brian E., P. Harik, and Stephen G. Clyman. 2000. The Generalizability of Scores for a Performance Assessment Scored with a Computer-Automated Scoring System. *Journal of Educational Measurement* 37:245–261.

- Clauser, Brian E., Melissa J. Margolis, Stephen G. Clyman, and Linette P. Ross. 1997. Development of Automated Scoring Algorithms for Complex Performance Assessments: A Comparison of Two Approaches. *Journal of Educational Measurement* 34:141–161.
- Clauser, Brian E., R. G. Subhiah, Ronald J. Nungester, D. R. Ripkey, Stephen G. Clyman, and D. McKinley. 1995. Scoring a Performance-Based Assessment by Modeling the Judgments of Experts. *Journal of Educational Measurement* 32:397–415.
- Clauser, Brian E., D. B. Swanson, and Stephen G. Clyman. 1999. A Comparison of the Generalizability of Scores Produced by Expert Raters and Automated Scoring Systems. *Applied Measurement in Education* 12:281–299.
- Cole, Jason C., and Anthony D. Lutkus. 1997. Score Comparisons of ACCUPLACER (Computer-Adaptive) and COMPANION (Paper) Reading Tests: Empirical Validation and School Policy. *Research in the Schools* 4 (2): 65–70.
- College Entrance Examination Board, ed. 1991. *Computerized Placement Tests: Background Readings*. ERIC Document Reproduction Service, ED 339 728.
- Conference on College Composition and Communication Committee (Kathleen Yancey, Chair). 2004. Position Statement on Teaching, Learning, and Assessing Writing in Digital Environments. *College Composition and Communication*, 55 (4): 785–789. www.ncte.org/groups/cccc/positions/115775.htm.
- Coombe, Christine A., ed. 1998. *Current Trends in English Language Testing: Conference Proceedings for CTELT 1997 and 1998*. Vol. 1. ERIC Document Reproduction Service, ED 428 574.
- Coombs, Don H. 1969. Review of *The Analysis of Essays by Computer*, by Ellis B. Page and Dieter H. Paulus. *Research in the Teaching of English* 3 (2): 222–228.
- Cooper, Charles R. 1974. Research Roundup: Oral and Written Composition: Henry Slotnick, A Model of Computer Essay Grading. *English Journal* 63: 103–104.
- Daigon, Arthur. 1966. Computer Grading of English Composition. *English Journal* 55 (1): 46–52.
- Dale, Robert, and Shona Douglas. 1997. Two Investigations into Intelligent Text Processing. In *The New Writing Environment: Writers at Work in a World of Technology*, edited by Mike Sharples and Thea van der Geest. Berlin: Springer-Verlag.
- Davies, Roy. 1989. The Creation of New Knowledge by Information Retrieval and Classification. *Journal of Documentation* 45 (4): 273–301.
- Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. 1990. Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science* 41:391–407.
- Dieterich, Daniel J. 1972. Composition Evaluation: Options and Advice. *English Journal* 61 (8): 1264–1271.
- Dobrin, David N. 1985. *Limitations on the Use of Computers in Composition*. ERIC Document Reproduction Service, ED 261 395.
- Drechsel, Joanne. 1999. Writing into Silence: Losing Voice with Writing Assessment Technology. *Teaching English in the Two Year College* 26:380–387.
- Elliot, Scott. 2003. IntelliMetric™: From Here to Validity. In Shermis and Burstein 2003.
- Felix, Uschi. 1993. Marking: A Pain in the Neck! The Computer to the Rescue. *Babel: Australia* 28 (3): 15–15.
- Finn, Patrick J. 1977. Computer-Aided Description of Mature Word Choices in Writing. In *Evaluating Writing: Describing, Measuring, Judging*, edited by Charles R. Cooper and Lee Odell. Urbana, IL: NCTE.
- Fitzgerald, Kathryn R. 1994. Computerized Scoring? A Question of Theory and Practice. *Journal of Basic Writing* 13 (2): 3–17.
- Foltz, Peter. 1990. Using Latent Semantic Indexing for Information Filtering. In *Proceedings of the Conference on Office Information Systems*, edited by R. B. Allen. New York: Association for Computing Machinery.
- . 1996. Latent Semantic Analysis for Text-Based Research. *Behavior Research Methods, Instruments, and Computers* 28:197–202.

- . 1998. Quantitative Approaches to Semantic Knowledge Representations. *Discourse Processes* 25 (2–3): 127–130.
- Foltz, Peter W., Darrell Laham, and Thomas K. Landauer. 1999. Automated Essay Scoring: Applications to Educational Technology. In *Ed Media 1999: World Conference on Educational Multimedia, Hypermedia, and Telecommunications, Seattle, Washington, USA, June 19–24, 1999*, edited by Betty Collis and Ron Oliver. Charlottesville, VA: Association for the Advancement of Computing in Education.
- Forehand, Garlie A. 1987. Development of a Computerized Diagnostic Testing Program. *Collegiate Microcomputer* 5 (1): 55–59.
- Forehand, Garlie A., and Myrtle W. Rice. 1987–1988. Diagnostic Assessment in Instruction. *Machine-Mediated Learning* 2 (4): 287–296.
- Hake, Rosemary. 1977. A Test to Teach to: Composition at Chicago State University. *ADE Bulletin* 52:31–37.
- Hartnett, Carolyn G. 1978. *Measuring Writing Skills*. ERIC Document Reproduction Service, ED 170 014.
- Hearst, Marti A. 2000. The Debate on Automated Essay Grading. *IEEE Intelligent Systems* 15 (5): 22–35. www.knowledge-technologies.com/papers/IEEEdedebate.pdf.
- Herrington, Anne, and Charles Moran. 2001. What Happens When Machines Read Our Students' Writing? *College English* 63 (4): 480–499.
- Hiller, Jack H. 1998. Applying Computerized Text Measurement Strategies from Project Essay Grade (PEG) to Military and Civilian Organizational Needs. ERIC Document Reproduction Service, ED 418 995.
- Hirschman, Lynette, Eric Breck, Marc Light, John D. Burger, and Lisa Ferro. 2000. Automated Grading of Short-Answer Tests. *IEEE Intelligent Systems* 15 (5): 31–35.
- Hock, Dennis J. 1990. A Correlational Study of Computer-Measured Traits and Holistic Assessment Ratings of First-Year College Students' Essays. Ph.D. diss., Indiana University of Pennsylvania.
- Huot, Brian. 1996. Computers and Assessment: Understanding Two Technologies. *Computers and Composition* 13 (2): 231–244.
- Jobst, Jack. 1984. Computer-Assisted Grading of Essays and Reports. *Computers and Composition* 1 (2): 5.
- . 1984. Computer-Assisted Grading: The Electronic Handbook. *Journal of Teaching Writing* 3 (2): 225–235.
- Johnson, Valen E. 1996. On Bayesian Analysis of Multirater Ordinal Data: An Application to Automated Essay Grading. *Journal of the American Statistical Association* 91 (433): 42–51.
- Jones, Brett D. 1999. Computer-Rated Essays in the English Composition Classroom. *Journal of Educational Computing Research* 20 (2): 169–187.
- Kaplan, Randy M., and Randy E. Bennett. 1994. *Using the Free-Response Scoring Tool to Automatically Score the Formulating-Hypothesis Item*. ETS Technical Report GRE No. 90-02b. Princeton, NJ: Educational Testing Service.
- Kaplan, Randy M., Susanne Wolff, Jill Burstein, Chi Lu, and Bruce Kaplan. 1998. *Scoring Essays Automatically Using Surface Features*. GRE Report No. 94-21P. Princeton, NJ: Educational Testing Service.
- Keith, Timothy Z. 2003. Validity of Automated Essay Scoring Systems. In Shermis and Burstein 2003.
- Kelly, P. Adam. 2001. *Computerized Scoring of Essays for Analytical Writing Assessments: Evaluating Score Validity*. ERIC Document Reproduction Service, ED 458 296.
- Klein, Davina C. D., Gregory K. W. K. Chung, Ellen Osmundson, Howard E. Herl, and Harold F. O'Neil, Jr. 2001. *Examining the Validity of Knowledge Mapping as a Measure of Elementary Students' Scientific Understanding*. CSE Technical Report No. 557. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Koether, Mary, and Esther Coke. 1973. *A Scheme for Text Analysis Using FORTRAN*. ERIC Document Reproduction Service, ED 074 152.

- Kukich, Karen. 2000. Beyond Automated Essay Scoring. *IEEE Intelligent Systems* 15 (5): 22–27.
- Landauer, Thomas K., and S. T. Dumais. 1997. A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review* 104:211–240.
- Landauer, Thomas K., Peter W. Foltz, and Darrell Laham. 2000. Introduction to Latent Semantic Analysis. *Discourse Processes* 25:259–284.
- Landauer, Thomas K., Darrell Laham, and Peter W. Foltz. 2000. The Intelligent Essay Assessor. *IEEE Intelligent Systems* 15:27–31.
- . 2003. Automated Scoring and Annotation of Essays with the Intelligent Essay Assessor™. In Shermis and Burstein 2003.
- Landauer, Thomas K., Darrell Laham, Bob Rehder, and M. E. Schreiner. 1997. How Well Can Passage Meaning Be Derived without Using Word Order? A Comparison of Latent Semantic Analysis and Humans. In *Proceedings of the Nineteenth Annual Meeting of the Cognitive Science Society: August 7–10, 1997, Stanford University, Stanford, CA*, edited by Michael G. Shafto and Pat Langley. Mahwah, NJ: Lawrence Erlbaum
- Larkey, Leah S. 1998. Automatic Essay Grading Using Text Categorization Techniques. In *SIGIR Ninety-Eight: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 24–28, 1998, Melbourne, Australia*, edited by W. Bruce Croft. New York: Association for Computing Machinery.
- Larkey, Leah S., and W. Bruce Croft. 2003. A Text Categorization Approach to Automated Essay Grading. In Shermis and Burstein 2003.
- Leacock, Claudia, and Martin Chodorow. 2003. Automated Grammatical Error Detection. In Shermis and Burstein 2003.
- Lee, Yong-Won. 2001. *The Essay Scoring and Scorer Reliability in TOEFL CBT*. ERIC Document Reproduction Service, ED 455 253.
- Lee, Yong-Won, Hunter Breland, and Eiji Muraki. 2002. *Comparability of TOEFL CBT Essay Prompts for Different Native Language Groups*. ERIC Document Reproduction Service, ED 464 963.
- Legg, Sue M., and Dianne C. Buhr. 1992. Computerized Adaptive Testing with Different Groups. *Educational Measurement: Issues and Practice* 11 (2): 23–27.
- Lemaire, Benoit, and Philippe Dessus. 2001. A System to Assess the Semantic Content of Student Essays. *Journal of Educational Computing Research* 24 (3): 305–320.
- Levy, Lynn B., and Kentner V. Fritz. *Status Report on the Computer Grading of Essays*. Counseling Center Reports, Vol. 5, No. 10. Madison: University of Wisconsin, Counseling Center. ERIC Document Reproduction Service, ED 069 759.
- Macrorie, Ken. 1969. Review of *The Analysis of Essays by Computer*, by Ellis B. Page and Dieter H. Paulus. *Research in the Teaching of English* 3 (2): 228–236.
- . 1970. Percival. In *Uptaught*. New York: Hayden.
- Manzo, Kathleen Kennedy. 2003. Essay Grading Goes Digital: But Critics Question the Use of Software to Assess Writing. *Education Week* 22 (35): 39–43.
- Martinez, M. E., and Randy E. Bennett. 1992. A Review of Automatically Scorable Constructed-Response Item Types for Large-Scale Assessment. *Applied Measurement in Education* 5 (2): 151–169.
- Matalene, Carolyn B., and Nancy Barendse. 1989. Transaction in Holistic Scoring: Using a Computer to Understand the Process. *Journal of Teaching Writing* 8 (2): 87–107.
- McCollum, Kelly. 1998. How a Computer Program Learns to Grade Essays. *Chronicle of Higher Education*, September 4: A37 A38.
- . 1999. Computers Will Help Grade Essays on Graduate Management Admission Test. *Chronicle of Higher Education*, January 29, A30.
- McCurry, Niki. 1992. *The Computerized Inventory of Developmental Writing Traits*. ERIC Document Reproduction Service, ED 357 375.
- McCurry, Niki, and Alan McCurry. 1992. Writing Assessment for the Twenty-First Century. *Computing Teacher* 19 (7): 35–37.
- Millman, Jason, and Ronald S. Westman. 1989. Computer-Assisted Writing of Achievement Test Items: Toward a Future Technology. *Journal of Educational Measurement* 26 (2): 177–190.

- Moon, Okhee. 1992. An Application of Computerized Adaptive Testing to the Test of English as a Foreign Language. Ph.D. diss., State University of New York at Albany.
- Morgan, Bradford A. 1984. Evaluating Student Papers with a Word Processor. *Research in Word Processing Newsletter* 2 (6): 1–6.
- Myers, Miles 2003. What Can Computers and AES Contribute to a K–12 Writing Program? In Shermis and Burstein 2003.
- Myford, Carol M., and Frederick Cline. 2002. Looking for Patterns in Disagreements: A Facets Analysis of Human Raters' and E-rater's Scores on Essays Written for the Graduate Management Admissions Test (GMAT). Paper presented at the annual meeting of the American Educational Research Association. www.ets.org/research/dload/AERA2002-myf.pdf.
- Mzumara, Howard R., Mark D. Shermis, and Jason M. Averitt. 1999. *Predictive Validity of the IUPUI Web-Based Placement Test Scores for Course Placement at IUPUI: 1998–1999*. Indianapolis: Indiana University Purdue University, Testing Center. www.assessment.iupui.edu/report/VAL1999.pdf.
- Mzumara, Howard R., Mark D. Shermis, and M. Fogel. 1999. *Predictive Validity of the IUPUI Web-Based Placement Test Scores for Course Placement at IUPUI: 1997–1998*. Indianapolis: Indiana University Purdue University, Testing Center.
- Nancarrow, Paula Reed, Donal Ross, and Lillian Bridwell-Bowles. 1984. *Word Processors and the Writing Process: An Annotated Bibliography*. Westport, CT: Greenwood.
- Ohlsson, Stellan. 1986. *Computer Simulation and its Impact on Educational Research and Practice*. ERIC Document Reproduction Service, ED 282 521.
- Ott, Christopher. 1999. Essay Questions: How Well Can Computers Judge Prose—And Would You Want One Grading Your Exam? *Salon.com*, May 25. www.salonmag.com/tech/feature/1999/05/25/computer_grading/.
- Page, Ellis B. 1966. Grading Essays by Computer: Progress Report. In *Proceedings of the Invitational Conference on Testing Problems, October 29, 1966, New York City*. Princeton, NJ: Educational Testing Service.
- . 1966. The Imminence of Grading Essays by Computer. *Phi Delta Kappan* 48:238–243.
- . 1967. Statistical and Linguistic Strategies in the Computer Grading of Essays. In *Proceedings of the Second International Conference on Computational Linguistics, Grenoble, France, August 24*. Yorktown Heights, NY: IBM Research Center.
- . 1968. Analyzing Student Essays by Computer. *International Review of Education* 14:210–225.
- . 1985. Computer Grading of Student Essays. In *International Encyclopedia of Educational Research: Research and Studies*, edited by Torsten Husen and T. Neville Postlethwaite. Oxford: Pergamon.
- . 2003. Project Essay Grade: PEG. In Shermis and Burstein.
- Page, Ellis B., G. A. Fisher, and M. A. Fisher. 1968. Project Essay Grade: A FORTRAN Program for Statistical Analysis of Prose. *British Journal of Mathematical and Statistical Psychology* 21:139.
- Page, Ellis B., and Dieter H. Paulus. 1968. *The Analysis of Essays by Computer*. Final Report of U.S. Office of Education Project No. 6-1318. Washington, DC: Department of Health, Education, and Welfare. ERIC Document Reproduction Service, ED 028 633.
- Page, Ellis B., and Nancy S. Petersen. 1995. The Computer Moves into Essay Grading: Updating the Ancient Test. *Phi Delta Kappan* 76:561–565.
- Page, Ellis B., John P. Poggio, and Timothy Z. Keith. 1997. Computer Analysis of Student Essays: Finding Trait Differences in Student Profile. Paper presented at the AERA/NCME Symposium on Grading Essays by Computer. ERIC Document Reproduction Service, ED 411 316.
- Page, Ellis B., P. I. Tillett, and Helen Beuck Ajay. 1989. Computer Measurement of Subject-Matter Tests: Past Research and Future Promises. In *Proceedings of the First Annual Meeting of the American Psychological Society*. Alexandria, VA: American Psychological Society.

- Palmer, John, Robert Williams, and Heinz Dreher. 2002. Automated Essay Grading System Applied to a First Year University Subject—Can We Do It Better? *InformingScience.org*. proceedings.informingscience.org/IS2002Proceedings/papers/Palme026Autom.pdf.
- Pedersen, Elray L. 1983. Computer-Assisted Evaluation of Student Papers: I Can Write Anything You Can Write—Faster And Better. *CALICO Journal* 1 (2): 39–42.
- Ponisciak, Steve, and Valen Johnson. 2003. Bayesian Analysis of Essay Grading. In Shermis and Burstein.
- Powers, Donald E., Jill C. Burstein, Martin S. Chodorow, Mary E. Fowles, and Karen Kukich. 2001. *Stumping E-rater: Challenging the Validity of Automated Essay Scoring*. GRE Board Research Report No. 98-08bP. Princeton, NJ: Educational Testing Service.
- . 2002. Comparing the Validity of Automated and Human Scoring of Essays. *Journal of Educational Computing Research* 26 (4): 407–425.
- Reid, Stephen, and Gilbert Findlay. 1986. Writer's Workbench Analysis of Holistically Scored Essays. *Computers and Composition* 3 (2): 6–32.
- Reitman, Walter R. 1962. Computer Models of Psychological Processes and Some Implication for the Theory and Practice of Writing. In *Needed Research in the Teaching of English: Proceedings of a Project English Research Conference, May 5–7, 1962*, edited by Erwin Ray Steinberg. Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.
- Roberts, Carl W., and Roel Popping. 1993. Computer-Supported Content Analysis: Some Recent Developments. *Social Science Computer Review* 11:283–291.
- Rounds, Jeanine C., Martha J. Kanter, and Marlene Blumin. 1987. Technology and Testing: What Is Around the Corner? In *Issues in Student Assessment*, edited by Dorothy Bray and Marcia J. Belcher. San Francisco: Jossey-Bass.
- Roy, Emil. 1990. A Decision Support System for Improving First-Year Writing Courses. *Computer-Assisted Composition Journal* 4 (3): 79–86.
- . 1993. Computerized Scoring of Placement Exams: A Validation. *Journal of Basic Writing* 12 (2): 41–54.
- . 1992. Evaluating Placement Exams with a Structured Decision System. *Computers and Composition* 9 (2): 71–86.
- Roy, Sandra, and Emil Roy. 1992. Direct Mail Letters: A Computerized Linkage between Style and Success. *Journal of Business and Technical Communication* 6 (2): 224–234.
- Rudner, Lawrence M. 2001. Bayesian Essay Test Scoring System—Betsy. ericae.net/betsy/.
- . 2001. Responding to Testing Needs in the 21st Century with an Old Tool, a Very Old Tool. In *Assessment: Issues and Challenges for the New Millennium*, edited by Garry Richard Walz and Jeanne Bleuer. Greensboro, NC: ERIC Clearinghouse on Counseling and Student Services.
- Rudner, Lawrence, and Phil Gagné. 2001. An Overview of Three Approaches to Scoring Written Essays by Computer. *Practical Assessment Research and Evaluation* 7(26). ericae.net/pare/getvn.asp?v=7&n=26
- Russell, Michael. 2000. *It's Time to Upgrade: Tests and Administration Procedures for the New Millennium*. ERIC Document Reproduction Service, ED 452 833.
- Russell, Michael, and Walt Haney. 2000. Bridging the Gap between Testing and Technology in Schools. *Education Policy Analysis Archives* 8 (19). epaa.asu.edu/epaa/v8n19.html.
- Salton, Gerald, and Anita Wong. 1976. On the Role of Words and Phrases in Automatic Text Analysis. *Computers and the Humanities* 10:69–87.
- Sampson, James P., Jr. 1998. *Using the Internet to Enhance Test Interpretation*. ERIC Document Reproduction Service, ED 426 329.
- Schachter, Ronald. 2003. A.I.: Something to Build on. *District Administration* 39 (July): 20–24.
- Schuemann, Cynthia M. 1998. An Investigation into the Comparable Effectiveness of Two Tests of Writing Administered to College-Level English as a Second Language Students: The Computerized Placement Test and Holistically Scored Essay Exams. Ph.D. diss., Florida International University, University Park.

- Segall, Michaela. 1991. Holistic Assessment of Writing. In *Project Synergy: Software Support for Underprepared Students*, edited by Kamala Anandam. Software Implementation Report, September. ERIC Document Reproduction Service, ED 345 803.
- Shermis, Mark D., and Fellicia D. Barrera, eds. 2002. *Automated Essay Scoring for Electronic Portfolios*. San Francisco: Wiley. [Special issue of *Assessment Update* 14 (4), 2002.]
- Shermis, Mark D., and Jill Burstein, eds. 2003. *Automated Essay Scoring: A Cross-Disciplinary Perspective*. Mahwah, NJ: Lawrence Erlbaum.
- Shermis, Mark D., and Kathryn E. Daniels. 2003. Norming and Scaling for Automated Essay Scoring. In Shermis and Burstein 2003.
- Shermis, Mark D., Chantal Mees Koch, Ellis B. Page, Timothy Z. Keith, and Susanne Harrington. 2002. Trait Ratings for Automated Essay Grading. *Educational and Psychological Measurement* 62 (1): 5–18.
- Shermis, Mark D., Howard Mzumara, Mike Brown, and Clo Lillig. 1997. *Computerized Adaptive Testing through the World Wide Web*. ERIC Document Reproduction Service, ED 414 536.
- Shukur, Zarina, Edmund Burke, and Eric Foxley. 1999. Automatically Grading Customer Confidence in a Formal Specification. *Journal of Computing in Higher Education* 11 (1): 86–119.
- Sirc, Geoffrey. 1989. Response in the Electronic Medium. In *Writing and Response: Theory, Practice, and Research*, edited by Chris M. Anson. Urbana, IL: NCTE.
- Sireci, Stephen G., and Saba Rizavi. 2000. *Comparing Computerized and Human Scoring of Students' Essays*. ERIC Document Reproduction Service, ED 463 324.
- Slotnick, Henry Barry. 1971. An Examination of the Computer Grading of Essays. Ph.D. diss., University of Illinois at Urbana-Champaign.
- . 1971. *An Examination of the Computer Grading of Essays*. ERIC Document Reproduction Service, ED 074 504.
- . 1972. Toward a Theory of Computer Essay Grading. *Journal of Educational Measurement* 9 (4): 253–263.
- . 1974. Computer Scoring of Formal Letters. *Journal of Business Communication* 11 (2): 11–19.
- Slotnick, Henry B., and John V. Knapp. 1971. Easy Grading by Computers: A Laboratory Phenomenon? *English Journal* 60 (1): 75–87.
- Slotnick, Henry B., John V. Knapp, and Rodney L. Bussell. 1971. Bits, Nybbles, Bytes: A View of Electronic Grading. *Journal of Business Communication* 8 (2): 35–52.
- Stansfield, Charles W., ed. 1986. *Technology and Language Testing: A Collection of Papers from the 7th Annual Colloquium on Language Testing Research, Princeton, New Jersey, April 6–9, 1985*. ERIC Document Reproduction Service, ED 365 142.
- Starr, Douglas. 1993. Paper Grading on Single Disc Preferable to Multi-Disc Way. *Community College Journalist* 21 (4): 4–6.
- Stephens, Derek. 2001. Use of Computer Assisted Assessment: Benefits to Students and Staff. *Education for Information* 19:265–275.
- Stewig, John Wren. 1984. The Key to Improving Education. *Education Digest* 49 (May): 10–13.
- Streeter, Lynn, Joseph Psotka, Darrell Laham, and Don MacCuish. 2004. The Credible Grading Machine: Automated Essay Scoring in the DOD [Department of Defense]. www.knowledge-technologies.com/papers/essayscoring.pdf.
- Takayoshi, Pamela. 1996. The Shape of Electronic Writing: Evaluating and Assessing Computer-Assisted Writing Processes and Products. *Computers and Composition* 13 (20): 245–258.
- Thompson, Clive. 1999. The Attack of the Incredible Grading Machine. *Lingua Franca* 9 (5). www.linguafranca.com/9907/nwo.html.
- United States Distance Learning Association. 2002. Underachieving Schools to Receive Online Development Tool. *USDLA Journal* 16 (July). www.usdla.org.

- Wainer, Howard, and Neil J. Dorans. 1990. *Computerized Adaptive Testing: A Primer*. Hillsdale, NJ: Lawrence Erlbaum.
- Weigle, Sara Cushing. 2002. *Assessing Writing*. Cambridge Language Assessment Series. Cambridge: Cambridge University Press.
- Whalen, Thomas E. 1971. *The Analysis of Essays by Computer: A Simulation of Teachers' Ratings*. ERIC Document Reproduction Service, ED 048 352.
- . 1971. *Assessment of Language Ability by Computer*. ERIC Document Reproduction Service, ED 052 206.
- Whithaus, Carl. 2004. The Development of Early Computer-Assisted Writing Instruction (1960–1978): The Double Logic of Media And Tools. *Computers and the Humanities* 38 (2): 149–162.
- Whittington, Dave, and Helen Hunt. 1999. Approaches to the Computerized Assessment of Free Text Responses. In *Proceedings of the Third Annual Computer Assisted Assessment Conference*. Loughborough, UK: Loughborough University. cvu.strath.ac.uk/dave/publications/caa99.html.
- Williams, Robert. 2001. Automated Essay Grading: An Evaluation of Four Conceptual Models. In *Expanding Horizons in Teaching and Learning: Proceedings of the 10th Annual Teaching Learning Forum, February 7–9, 2001*, edited by A. Herrmann and M. M. Kulski. Perth, Australia: Curtin University of Technology. cea.curtin.edu.au/tlf/tlf2001/williams.html.
- Williamson, David M., Isaac I. Bejar, and A. S. Hope. 1999. “Mental Model” Comparison of Automated and Human Scoring. *Journal of Educational Measurement* 36:158–184.
- Williamson, Michael. 2004. Validity of Automated Scoring: Prologue for a Continuing Discussion of Machine Scoring Student Writing. *Journal of Writing Assessment* 1 (2): 85–104.
- Wirth, Susan Katherine. 1972. A Paraphrasing Strategy in the Content Grading of Essay Examinations by Computer, Ph.D. diss., University of Connecticut, Storrs.
- Wonnberger, Carl G. 1955. Judging Compositions—Machine Method. *English Journal* 44 (8): 473–475.
- Woods, Elinor M. 1970. Recent Applications of Computer Technology to School Testing Programs. *Review of Educational Research* 40 (4): 525–539.
- Wresch, William. 1988. Six Directions for Computer Analysis of Student Writing. *Computing Teacher* 15 (7): 13–16.
- . 1993. The Imminence of Grading Essays by Computer—25 Years Later. *Computers and Composition* 10 (2): 45–58.
- Wresch, William, and Helen Schwartz. 1988. The Uses of Computers in the Analysis and Assessment of Writing [conference presentation synopsis]. *Notes from the National Testing Network in Writing* 8:26–27.
- Yang, Yongwei, Chad W. Buckendahl, Piotr J. Juskiewicz, and Dennison S. Bhola. 2002. A Review of Strategies for Validating Computer Automated Scoring. *Applied Measurement in Education* 15 (4): 391–412.