**AUTOMATED ESSAY SCORING AND ELECTRONIC FEEDBACK:**

**SELECTED REFERENCES**

**(Last updated 18 December 2016)**

Attali, Y. (2004). *Exploring the feedback and revision features of Criterion*. Paper presented at the National Council on Measurement in Education (NCME), San Diego, CA.

Attali, Y. (2007). *Construct validity of e-rater in scoring TOEFL essays* (Research Report RR-07-21). Princeton, NJ: Educational Testing Service.

Attali, Y. (2011). *Automated subscores for TOEFL iBT independent essays* (Research Report RR-11-39). Princeton, NJ: Educational Testing Service.

Attali, Y., Bridgeman, B., & Trapani, C. (2010). Performance of a generic approach in automated essay scoring. *Journal of Technology, Learning, and Assessment, 10*(3), 4-16.

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V. 2. *The Journal of Technology, Learning and Assessment, 4*(3), 3-30.

Attali, Y., & Sinharay, S. (2015). *Automated trait scores for TOEFL writing tasks* (Research Report RR-15-14). Princeton, NJ: Educational Testing Service.

Ben-Simon, A., & Bennett, R. E. (2007). Toward more substantively meaningful automated essay scoring. *Journal of Technology, Learning, and Assessment, 6*(1), 1-47.

Bennett, R. E. (2004). *Moving the field forward: Some thoughts on validity and automated scoring*. Princeton, NJ: Lawrence Erlbaum Associates.

Bennett, R. E., & Bejar, I. I. (1998). Validity and automated scoring: It’s not only the scoring. *Educational Measurement: Issues and Practice, 17*(4), 9-17.

Bereiter, C. (2003). Foreword. In M. D. Shermis & J. C. Burstein (Eds.), *Automated essay scoring: A cross disciplinary perspective* (pp. vii- ix). Mahwah, NJ: Lawrence Erlbaum.

Bernstein, J., De Jong, J., Pisoni, D., & Townshend, B. (2000). Two experiments on automatic scoring of spoken language proficiency. In P. Delcloque (Ed.), *Proceedings in InSTIL2000* (pp. 57-61). Dundee, Scotland: University of Abertay.

Bridgeman, B., Trapani, C, & Attali, Y. (2012). Comparison of human and machine scoring of essays: Differences by gender, ethnicity, and country. *Applied Measurement in Education, 25*(1), 27-40.

Bridgeman, B., Trapani, C., & Attali, Y. (2009, April). *Considering fairness and validity in evaluating automated scoring.* Paper presented at the annual meeting of the National Council on Measurement in Education (NCME), San Diego, CA. Retrieved from http://www.ets.org/Media/Conferences\_and\_Events/AERA\_2009\_pdfs/AERA\_NCME\_2009\_Bridgeman.pdf

Brill, E., & Mooney, R. (1997). An overview of empirical natural language processing. *AI Magazine, 18*(4), 13-24.

Burstein, J. (2003). The *E-rater*® scoring engine: Automated essay scoring with natural language processing. In M. D. Shermis & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 113-121). Mahwah, NJ: Lawrence Erlbaum Associates.

Burstein, J., & Chodorow, M. (1999). Automated essay scoring for nonnative English speakers. Proceedings from the *ACL99 Workshop on Computer-Mediated Language Assessment and Evaluation of Natural Language Processing*, College Park, MD.

Burstein, J., Chodorow, M., & Leacock, C. (2003, August). Criterion: Online essay

evaluation: An application for automated evaluation of student essays.Proceedings

from the *15th Annual Conference on Innovative Applications of Artificial Intelligence,*

Acapulco, Mexico. Retrieved from http://www.ets.org/Media/Research/pdf/erater\_iaai03\_burstein.pdf

Burstein, J., Kukich, K., Wolff, S., Lu, C., & Chodorow, M. (1998, April). *Computer analysis of essays*. Proceedings from the *NCME Symposium on Automated Scoring*, Montreal, Canada.

Burstein, J., Leacock, C., & Swartz, R. (2001). Automated evaluation of essays and short answers*.* Proceedings from the *5th International Computer Assisted Assessment Conference (CAA 01)*, Loughborough University, UK.

Burstein, J., & Marcu, D. (2000). Benefits of modularity in an automated essay scoring system. In Proceedings from the *Workshop on Using Toolsets and Architectures to Build NLP Systems, 18th International Conference on Computational Linguistics,* Luxembourg. Retrieved from ERIC database. (ED447168)

Burstein, J., & Wolska, M. (2003). *Toward evaluation of writing style: Finding overly repetitive word use in student essays*. In *EACL ’03: Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics* (pp. 35-42). Morristown, NJ: Association for Computational Linguistics. Retrieved from <http://www.aclweb.org/anthology-new/E/E03/E03-1003.pdf>

Chapelle, C. (2001). *Computer applications in second language acquisition*. Cambridge, UK: Cam­bridge University Press.

Chen, C., & Cheng, W. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology, 12*(2), 94-112.

Cheville, J. (2004). Automated scoring technologies and the rising influence of error. *English Journal, 93*(4), 47-52.

Chodorow, M., & Burstein, J. (2004). *Beyond essay length: Evaluating e-rater’s performance on TOEFL essays* (Research report No. 73)*.* Princeton, NJ: Educational Testing Service.

Chung, K. W. K., & O’Neil, H. F. (1997). Methodological approaches to online scoring of essays.Retrieved from http://www.cse.ucla.edu/products/reports/tech461.pdf

Cizek, G. J., & Page, B. A. (2003). The concept of reliability in the context of automated essay scoring. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 125-145). Mahwah, NJ: Lawrence Erlbaum.

Conference on College Composition and Communication. (2006) .*Writing assessment: A position statement*. Retrieved from http://www.ncte.org/cccc/resources/positions/writingassessment

Cooze, M. (2011). Assessing writing tests on scoris®: The introduction of online marking. *Cambridge ESOL Research Notes*, *43,* 12-15.

Cotos, E. (2009). Designing an intelligent discourse evaluation tool: Theoretical, empirical, and technological considerations. In C. A. Chapelle, G. H. Jun, & I. Katz (Eds.), *Developing and evaluating language learning materials* (pp. 103‐127). Ames, IA: Iowa State University.

Cotos, E. (2011). Potential of automated writing evaluation feedback. *CALICO Journal, 28*(2), 420-459.

Cotos, E., & Pendar, N. (2008). Automated diagnostics writing tests: Why? How? In C. A. Chapelle, Y.-R. Chung, & J. Xu (Eds.), *Towards adaptive CALL: Natural language processing for diagnostic language assessment* (pp. 65-81). Ames, IA: Iowa State University.

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing, 18*, 7-24.

Dikli, S. (2006) . An overview of automated scoring of essays. *Journal of Technology, Learning, and Assessment, 5*(1), 1-35.

Ebyary, K., &Windeatt, S. (2010). The impact of computer-based feedback on students’ written work. *International Journal of English Studies, 10*(2), 121-142.

Educational Testing Service (ETS). (n.d.). *E-rater*. Retrieved from http://www.ets.org/erater/about

Educational Testing Service (ETS). (n.d.). *Criterion.* Retrieved from

http://www.ets.org/criterion/ell/about/faq

Elliot, S. (2000). *A study of expert scoring and IntelliMetric scoring accuracy*

*for dimensional scoring of Grade 11 student writing responses* (RB- 397). Newtown,

PA: Vantage Learning.

Elliot, S. (2000). *A true score study of IntelliMetric accuracy for holistic and dimensional scoring of college entry-level writing program* (RB-407). Newtown, PA: Vantage Learning.

Elliot, S. (2001). *About IntelliMetric* (PB-540). Newtown, PA: Vantage Learning.

Elliot, S. (2001). *Applying IntelliMetric Technology to the scoring of 3rd and 8th grade standardized writing assessments* (RB-524). Newtown, PA: Vantage Learning.

Elliot, S. (2002). *A study of expert scoring, standard human scoring and IntelliMetric scoring accuracy for statewide eighth grade writing responses* (RB-726). Newtown, PA: Vantage Learning.

Elliott, S. (2002). *IntelliMetric Analytic Scoring practitioner’s guide to scientifically based*

*MY Access research* (RB- 759). Newtown, PA: Vantage Learning.

Elliot, S. (2003). *A true score study of 11th grade student writing responses using IntelliMetric Version 9.0* (RB-786). Newtown, PA: Vantage Learning.

Elliot, S. (2003). *Assessing the accuracy of IntelliMetric for scoring a district- wide writing assessment* (RB-806). Newtown, PA: Vantage Learning.

Elliot, S. (2003). *How does IntelliMetric score essay responses?* (RB-929). Newtown, PA: Vantage Learning.

Elliot, S. (2003). IntelliMetric: From here to validity. In M. D. Shermis & Jill C. Burstein (Eds.), *Automated essay scoring: A cross disciplinary approach*. Mahwah, NJ: Lawrence Erlbaum.

Enright, M., & Quinlan, M. (2010). Complementing human judgment of essays written by English language learners with e-rater® scoring. *Language Testing, 27*(3), 317-334.

Ericsson, P. F. (2006). The meaning of meaning. In P. F. Ericsson & R. Haswell (Eds.), *Machine scoring of human essays: Truth of consequences* (pp. 28-37). Logan, UT: Utah State University Press.

Foltz, P. W. (1996). Latent Semantic Analysis for text-based research. *Behavior Research Methods, Instruments and Computers, 28*(2), 197-202.

Foltz, P. W., Laham, D., & Landauer, T. K. (1999). Automated essay scoring: Applications to educational technology. Proceedings from *EdMedia '99*. Retrieved from http://www-psych.nmsu.edu/ ~pfoltz/reprints/Edmedia99.html

Grimes, D., & Warschauer, M. (2010). [Utility in a fallible tool: A multi-site case study of automated writing evaluation](http://escholarship.bc.edu/jtla/vol8/6/). *Journal of Technology, Language, and Assessment, 8*(6), 1-43.

Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing, 18*(3), 218-238.

Haberman, S. J. (2011). *Use of e-rater in scoring of the TOEFL iBT writing test* (Research Memorandum RM-11-25). Princeton, NJ: Educational Testing Service.

Hamp-Lyons, L. (2001). Fourth generation writing assessment. In T. Silva & P. K. Matsuda (Eds.), *On Second Language Writing* (pp. 117-125). Mahwah, NJ: Lawrence Erlbaum.

Higgins, D., Ramineni, C., & Zechner, K. (2015). The use of learner corpora for automated scoring of written and spoken responses. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 587-586)*.* Cambridge, UK: Cambridge University Press.

Hyland, F. (1998). The impact of teacher written feedback on individual writers. *Journal of Second Language Writing, 7*(3), 255-286.

Hyland, K., & Hyland, F. (Eds.). (2006). *Feedback in second language writing: Contexts and issues.* New York, NY: Cambridge University Press.

James, C. L. (2006). Validating a computerized scoring system for assessing writing and placing students in composition courses. *Assessing Writing*, *11*, 167-178.

Kakkonen, T., & Sutinen, E. (2008). Evaluation criteria for automated assessment systems-There is much more than just the correlation. In *Proceedings of the 16th International Conference on Computers in Education*, Taipei, Taiwan.

Kellogg, R. T., Whiteford, A. P., & Quinlan, T. (2010). Does automated feedback help students learn to write? *Journal of Educational Computing Research*, *42*(2), 173-196.

Kukich, K. (2000, September/October). Beyond automated essay scoring. *IEEE Intelligent systems:* *The debate on automated essay grading*, 22-27.

Lai, G., Calandra, B. (2010). Examining the effects of computer-based scaffolds on novice teachers’ reflective journal writing. *Education Technology Research and Development*, *58,* 421-437.

Landauer, T. K., Foltz, P. W., & Laham, D. (2004). *What is LSA?* Retrieved from <http://lsa.colorado.edu/whatis.html>

Landauer, T. K., Laham, D., & Foltz, P. W. (2000, September/ October). The Intelligent Essay Assessor. *IEEE Intelligent systems,* *The debate on automated essay grading*, 27- 31. Retrieved from

Landauer, T. K., Laham, D., & Foltz, P. W. (2003). Automatic essay assessment. *Assessment in Education,* *10*(3), 295-308.

Landauer, T. K., Laham, D., & Foltz, P. W. (2003). Automated essay scoring: A cross disciplinary perspective. In M. D. Shermis & J. C. Burstein (Eds.), *Automated Essay Scoring and annotation of essays with the Intelligent Essay Assessor* (pp. 87-112). Mahwah, NJ: Lawrence Erlbaum.

Landauer, T. K., Laham, D., Rehder, B. & Schreiner, M. E. (1997). How well can passage meaning be derived without using word order? A comparison of Latent Semantic Analysis and humans. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society* (pp. 412-417). Mahwah, NJ: Lawrence Erlbaum.

Lee, C., Wong, K., Cheung, W., & Lee, F. (2008). Web-based essay critiquing system and EFL students’ writing: A quantitative and qualitative investigation. *Computer Assisted Language Learning,* *22*(1), 57-72.

Lemaire, B., & Dessus, P. (2001). A system to assess the semantic content of student essays. *Educational Computing Research, 24*(3), 305-306.

Loukina, A., & Cahill, A. (2016). Automated scoring across different modalities. *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications,* 130-135.

Monaghan, W., & Bridgeman, B. (2005, April). E-rater as a quality control on human scorers. *ETS RD Connections*. Retrieved from <http://www.ets.org/Media/Research/pdf/RD_Connections2.pdf>

Murray, B. (1998). The latest techno tool: Essay grading computers. *American Psychological Association Monitor, 8*(29), 43.

Myers, M. (2003). What can computers and AES contribute to a K–12 writing program? In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 3-20). Mahwah, NJ: Lawrence Erlbaum.

Nagata, N. (1993). Intelligent computer feedback for second language instruction. *The Modern Language Journal, 77*(3), 330-339.

Nichols, P. D. (2004, April). *Evidence for the interpretation and use of scores from an Automated Essay Scorer*. Paper presented at the annual meeting of the American Educational Research Association (AERA), San Diego, CA.

Page, E. B. (1967). The imminence of grading essays by computer. *Phi Delta Kappan, 48,* 238-243.

Page, E. B. (1994). Computer grading of student prose, using modern concepts and software. *Journal of Experimental Education, 62*, 127-142.

Page, E. B. (2003). Project Essay Grade: PEG*.* In M. D. Shermis & J. Burstein (Eds.).

*Automated essay scoring: A cross-disciplinary perspective* (pp. 43-54)*.* Mahwah, NJ:

Lawrence Erlbaum.

Pendar, N., & Cotos, E. (2008). *Automatic identification of discourse moves in scientific article introductions.*Presentation at the 46th annual meeting of the Association for Computational Linguistics (ACL) with the Human Language Technology Conference (HLT) of the North American Chapter*,* Columbus, Ohio.

Phillips, S. (2007). *Automated essay scoring: A literature review*. Kelowna, B.C.: Society for the Advancement of Excellence in Education.

Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M. E., & Kukich, K. (2000). *Comparing the validity of automated and human essay scoring* (ETS RR-00-10). Princeton, NJ: ETS.

Powers, D., Burstein, J., Chodorow, M., Fowles, M., & Kukich, K. (2001). *Stumping E-rater: Challenging the validity of automated essay scoring* (Research Report 01-03). Princeton, NJ: Educational Testing Service.

Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M. E., Kukich, K., & Board, G. R. E. (2000). *Comparing the validity of automated and human essay scoring* (Research Report 00-10). Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/Media/Research/pdf/RR-00-10.pdf

Psotka, J., & Streeter, L.(n.d.). *Automatically critiquing writing for army educational settings*. Retrieved from <http://www.hqda.army.mil/ari/pdf/critiquing_writing.pdf>

Ramineni, C., Trapani, C. S., Williamson, D. M., Davey, T., & Bridgeman, B. (2012). *Evaluation of the e-rater scoring engine for the TOEFL independent and integrated prompts* (Research Report RR-12-06). Princeton, NJ: Educational Testing Service.

Riedel, E., Dexter, S., Scharber, C., & Doering, A. (2005, April). *Experimental evidence on the effectiveness of automated essay scoring in teacher education cases.* Paper presented at the 86th annual meeting of the American Educational Research Association, Montreal, Canada.

Rock, J. (2007). *The impact of short-term use of Criterion on writing skills in ninth grade* (Research Report 07-07). Princeton, NJ: Educational Testing Service.

Rudner, L. & Gagne, P. (2001). *An overview of three approaches to scoring written essays by computer.* ERIC Digest, ERIC Clearinghouse on Assessment and Evaluation. (ERIC Document Reproduction Service No. ED458290).

Rudner, L., Garcia, V., & Welch, C. (2005). *An evaluation of Intellimetric™ essay scoring system using responses to GMAT® AWA prompts* (GMAC Research Report 05-08).

McLean, VA: Graduate Management Admission Council.

Rudner, L. M., Garcia, V., & Welch, C. (2006). An evaluation of IntelliMetric™ essay scoring system. *The Journal of Technology, Learning and Assessment, 4*(4), 1-21.

Rudner, L. M., & Liang, T. (2002). Automated essay scoring using Bayes’ theorem. *The Journal of Technology, Learning, and Assessment, 1*(2).

Sadrolmomaleki, V. (2009). Electronic feedback and teaching writing. *International Journal of Arts and Sciences, 3*(5), 36-48.

Salem, A. B. M. (2000). *The potential role of artificial intelligence technology in education.* Retrieved from ERIC database. (ED 477318)

Shermis, M., & Barrera, F. (2002). *Exit assessments: Evaluating writing ability through Automated Essay Scoring.* Retrieved from ERIC database. (ED 464950)

Shermis, M. D., & Burstein, J. (2003). *Automated Essay Scoring: A cross disciplinary perspective*. Mahwah, NJ: Lawrence Erlbaum.

Shermis, M., Burstein, J., Brew, C., Higgins. D., & Zechner, K. (2015). Recent innovations in machine scoring of student and test taker written and spoken responses. In S. Lane, M. Raymond, & T. Haladyna (Eds.), *Handbook of test development* (pp. 335-354).  New York, NY: Routledge.

Shermis, M., Burstein, J., Higgins, D., & Zechner, K. (2010). Automated essay scoring: Writing assessment and instruction. In P. Peterson, E. Baker, and B. McGaw (Eds.), *International encyclopedia of education* (3rd ed., pp. 20-26). Oxford, UK: Elsevier.

Shermis, M. D., Raymat, M. V., & Barrera, F. (2003). *Assessing writing through the curriculum with Automated Essay Scoring.* Retrieved from ERIC database. (ED477929)

Sireci, S. G., & Rizavi, S. (1999). *Comparing computerized and human scoring of students’ essays.* Retrieved from ERIC database. (ED463324).

Streeter, L., Psotka, J., Laham, D., & MacCuish, D. (2004). The credible grading machine: Essay scoring in the DOD [Department of Defense]. Retrieved from <http://www.k-a-t.com/papers/essayscoring.pdf>

Valenti, S., Neri, F., & Cucchiarelli, A. (2003). An overview of current research on automated essay grading. *Journal of Information Technology Education, 2*, 319-330.

Vantage Learning. (n.d.). Retrieved from <http://www.vantagelearning.com>

Vantage Learning. (1999). *Construct validity of IntelliMetric with International Assessment* (RB-323). Newtown, PA: Vantage Learning.

Vantage Learning. (2000). *A study of expert scoring and IntelliMetric scoring accuracy for dimensional scoring of Grade 11 student writing responses* (RB-397). Newtown, PA: Vantage Learning.

Vantage Learning. (2000). *A true score study of IntelliMetric accuracy for holistic and dimensional scoring of college entry-level writing program* (RB-407). Newtown, PA:Vantage Learning.

Vantage Learning. (2001). *About IntelliMetric* (PB-540). Newtown, PA: Vantage Learning.

Vantage Learning. (2001). *Applying IntelliMetric Technology to the scoring of 3rd and 8th grade standardized writing assessments* (RB-524). Newtown, PA: Vantage Learning.

Vantage Learning. (2002). *A study of expert scoring, standard human scoring and IntelliMetric scoring accuracy for statewide eighth grade writing responses* (RB-726). Newtown, PA: Vantage Learning.

Vantage Learning. (2003). *Assessing the accuracy of IntelliMetric for scoring a district-wide writing assessment* (RB-806). Newtown, PA: Vantage Learning.

Vantage Learning. (2003). *How does IntelliMetric score essay responses?* (RB-929). Newtown, PA: Vantage Learning.

Vantage Learning. (2003). *A true score study of 11th grade student writing responses using IntelliMetric Version 9.0* (RB-786). Newtown, PA: Vantage Learning.

Wang, F., & Wang, S. (2012). A comparative study on the influence of automated evaluation system and teacher grading on students’ English writing. *Procedia Engineering*, *29*, 993- 997.

Wang, J., & Brown, M. (2007). Automated essay scoring versus human scoring: A comparative study. *The Journal of Technology Learning and Assessment, 6*(2), 1-29.

Ware, P., & Warschauer, M. (2006). [Electronic feedback and second language writing](http://gse.uci.edu/person/warschauer_m/docs/feedback.pdf). In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 105-122). New York, NY: Cambridge University Press.

Warschauer, M ., & Grimes, D. (2008). [Automated writing assessment in the classroom](http://gse.uci.edu/person/warschauer_m/docs/awe-pedagogies.pdf). *Pedagogies, 3*(1), 52-67.

Warschauer, M., & Ware, P. (2006). [Automated writing evaluation: Defining the classroom research agenda](http://gse.uci.edu/person/warschauer_m/docs/AWE.pdf). *Language Teaching Research, 10*(2), 1-24.

Warschauer, M., & Grimes, D. (2010).Utility in a fallible tool: A multi-site case study of automated writing evaluation. *The Journal of Technology, Learning, and Assessment, 8*(6), 1-41.

Weigle, S. C. (2010). Validation of automated scores of TOEFL iBT tasks against non-test indicators of writing ability. *Language Testing*, *27*(3), 335-353.

Weigle, S. C. (2011). *Validation of automated scores of TOEFL iBT tasks against nontest indicators of writing ability* (TOEFL iBT Research Report No. 15). Princeton, NJ: Educational Testing Service.

Williamson, D., Xi, X., & Breyer, J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement, Issues and Practice, 31*(1), 2-13.

Xi, X. (2010). Automated scoring and feedback systems: Where are we and where are we heading? *Language Testing,* *27*(3), 291-300.

Zechner, K., Higgins, D., Xi, X., & Williamson, D. M. (2009). Automatic scoring of non-native spontaneous speech in tests of spoken English. *Speech Communication, 51*(10), 883-895.