References

Generation of Exercises

— Computer-based Testing and Question Generation —

Duvall, K. Improving Your Test Questions. [Online; visited May 26, 2008]. Center for Teaching Excellence, University of Illinois at Urbana-Champaign. http://www.oir.uiuc.edu/dme/exams/ITQ.html.

McKenna, C. and Bull, J. (1999). Designing effective objective test questions: an introductory workshop. [Online; visited May 26, 2008]. CAA Centre, Loughborough University, http://caacentre.lboro.ac.uk/dldocs/otghdout.pdf.

— Multiple-choice Questions —

Brown, J. C., Frishkoff, G. A., and Eskenazi, M. (2005). Automatic question generation for vocabulary assessment. In *HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 819–826, Morristown, NJ, USA. Association for Computational Linguistics.

Heilman, M. and Eskenazi, M. (2007). Application of Automatic Thesaurus Extraction for Computer Generation of Vocabulary Questions. In *Proceedings of Speech and Language Technology in Education (SLaTE2007)*, pages 65–68.

Karamanis, N., Ha, L. A., and Mitkov, R. (2006). Generating Multiple-Choice Test Items from Medical Text: A Pilot Study. In *Proceedings of the Fourth International Natural Language Generation Conference*, pages 111–113, Sydney, Australia. Association for Computational Linguistics.

Mitkov, R., Ha, L. A., and Karamanis, N. (2006). A computer-aided environment for generating multiple-choice test items. *Natural Language Engineering*, 12(2):177–194.

— Fill-in-the-blank Questions —

Aldabe, I., de Lacalle, M. L., Maritxalar, M., Martinez, E., and Uria, L. (2006). ArikIturri: An Automatic Question Generator Based on Corpora and NLP Techniques. In Ikeda, M., Ashley, K. D., and Chan, T.-W., editors, *Intelligent Tutoring Systems*, volume 4053 of *Lecture Notes in Computer Science*, pages 584–594. Springer.

Coniam, D. (1997). A Preliminary Inquiry Into Using Corpus Word Frequency Data in the Automatic Generation of English Language Cloze Tests. *CALICO Journal*, 14:15–33.

— Multiple-choice Cloze Questions —

Chen, C.-Y., Liou, H.-C., and Chang, J. S. (2006). FAST: an automatic generation system for grammar tests. In *Proceedings of the COLING/ACL Interactive presentation sessions*, pages 1–4, Morristown, NJ, USA. Association for Computational Linguistics.

Hoshino, A. and Hiroshi, N. (2005). A Real-Time Multiple-Choice Question Generation For Language Testing: A Preliminary Study. In *Proceedings of the Second Workshop on Building Educational Applications Using NLP*, pages 17–20, Ann Arbor, Michigan. Association for Computational Linguistics.

Lee, J. and Seneff, S. (2007). Automatic Generation of Cloze Items for Prepositions. In *Proceedings of INTERSPEECH 2007*, pages 2173–2176, Antwerp, Belgium.

Liu, C.-L., Wang, C.-H., Gao, Z.-M., and Huang, S.-M. (2005). Applications of Lexical Information for Algorithmically Composing Multiple-Choice Cloze Items. In *Proceedings of the Second Workshop on Building Educational Applications Using NLP*, pages 1–8, Ann Arbor, Michigan. Association for Computational Linguistics.

Smith, S., Sommers, S., and Kilgarriff, A. (2008). Learning words right with the Sketch Engine and WebBootCat: Automatic cloze generation from corpora and the web. In *Proceedings of the Conference of English Teaching and Learning in R.O.C.*

Sumita, E., Sugaya, F., and Yamamoto, S. (2005). Measuring Non-native Speakers' Proficiency of English by Using a Test with Automatically-Generated Fill-in-the-Blank Questions. In *Proceedings of the Second Workshop on Building Educational Applications Using NLP*, pages 61–68, Ann Arbor, Michigan. Association for Computational Linguistics.

— **Matching Test Items** —

Brown, J. C., Frishkoff, G. A., and Eskenazi, M. (2005). Automatic question generation for vocabulary assessment. In *HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 819–826, Morristown, NJ, USA. Association for Computational Linguistics.

— **Error Correction Questions** —

Chen, C.-Y., Liou, H.-C., and Chang, J. S. (2006). FAST: an automatic generation system for grammar tests. In *Proceedings of the COLING/ACL Interactive presentation sessions*, pages 1–4, Morristown, NJ, USA. Association for Computational Linguistics.

— **Item Analysis** —

Zurawski, R. M. (1998). Making the Most of Exams: Procedures for Item Analysis. *The National Teaching & Learning FORUM*, 7(6):1–4.

Assessment of Learner-Generated Discourse

— **Essay Scoring** —

Attali, Y. and Burstein, J. (2006). Automated Essay Scoring With e-rater® V.2. *Journal of Technology, Learning and Assessment*, 4(3).

Breland, H. M., Jones, R. J., and Jenkins, L. (1994). The College Board vocabulary study. Technical report, College Board Report No. 94–4, New York: College Entrance Examination Board.

Burstein, J., Kukich, K., Wolff, S., Lu, C., Chodorow, M., Braden-Harder, L., and Harris, M. D. (1998). Automated scoring using a hybrid feature identification technique. In *Proceedings of the 17th international conference on Computational linguistics*, pages 206–210, Morristown, NJ, USA. Association for Computational Linguistics.

Burstein, J., Marcu, D., and Knight, K. (2003). Finding the WRITE Stuff: Automatic Identification of Discourse Structure in Student Essays. *IEEE Intelligent Systems*, 18(1):32–39.

Burstein, J. and 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*, pages 35–42, Morristown, NJ, USA. Association for Computational Linguistics.

Elliot, S. M. (2001). IntelliMetric: from here to validity. In *Paper presented at the annual meeting of the American Educational Research Association*, Seattle, WA.

Hearst, M. A. (2000). The Debate on Automated Essay Grading. *IEEE Intelligent Systems*, 15(5):22–37.

Higgins, D., Burstein, J., and Attali, Y. (2006). Identifying off-topic student essays without topic-specific training data. *Natural Language Engineering*, 12(2):145–159.

Landauer, T. K., Laham, D., and Foltz, P. (1998). Learning Human-like Knowledge by Singular Value Decomposition: A Progress Report. *Advances in Neural Information Processing Systems*, 10:45–51.

Page, E. B. (1966). The imminence of grading essays by computer. *Phi Delta Kappan*, 47:238–243.

Page, E. B. (1994). Computer Grading of Student Prose, Using Modern Concepts and Software. *Journal of Experimental Education*, 62:127–142.

Yang, Y., Buckendahl, C. W., Juszkiewicz, P. J., and Bhola, D. S. (2002). A Review of Strategies for Validating Computer-Automated Scoring. *Applied Measurement in Education*, 15(4):391–412.

<center>— **Plagiarism** —</center>

Clough, P. (2000). Plagiarism in Natural and Programming Languages: an Overview of Current Tools and Technologies. Technical report, Internal Report CS-00-05, University of Sheffield.

Clough, P. (2003). Old and new challenges in automatic plagiarism detection. Technical report, National UK Plagiarism Advisory Service.

Martin, B. (1994). Plagiarism: a misplaced emphasis. *Journal of Information Ethics*, 3(2):36–47.

<center>— **Short Answer Assessment** —</center>

Bailey, S. and Meurers, D. (2008). Diagnosing Meaning Errors in Short Answers to Reading Comprehension Questions. In *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications*, pages 107–115, Columbus, Ohio. Association for Computational Linguistics.

Leacock, C. (2004). Scoring free-responses automatically: A case study of a large-scale assessment. *Examens*, 1(3).

Leacock, C. and Chodorow, M. (2003). c-rater: Scoring of short-answer questions. *Computers and the Humanities*, 37:389–405.

Sandene, B., Horkay, N., Bennett, R. E., Allen, N., Braswell, J., Kaplan, B., , and Oranje, A. (2005). Online Assessment in Mathematics and Writing: Reports From the NAEP Technology-Based Assessment Project, Research and Development Series. Technical report, National Assessment of Educational Progresss.

<center>— **Speech Assessment** —</center>

Bernstein, J. (1999). PhonePass testing: Structure and construct. Technical report, Ordinate Corporation.

Bernstein, J., DeJong, J., Pisoni, D., and Townshend, B. (2000). Two experiments on automatic scoring of spoken language proficiency. In *Proceedings of InSTIL2000*.

Zechner, K. and Bejar, I. I. (2006). Towards automatic scoring of non-native spontaneous speech. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pages 216–223.

Zechner, K., Higgins, D., and Xi, X. (2007). SpeechRaterTM: A Construct-Driven Approach to Scoring Spontaneous Non-Native Speech. In *roceedings of the 2007 Workshop of the International Speech Communication Association (ISCA) Special Interest Group on Speech and Language Technology in Education (SLaTE2007)*.

Zechner, K. and Xi, X. (2008). Towards Automatic Scoring of a Test of Spoken Language with Heterogeneous Task Types. In *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications*, pages 98–106, Columbus, Ohio. Association for Computational Linguistics.

Reading and Writing Assistance

<center>— **Text Readability** —</center>

Brown, J. and Eskenazi, M. (2004). Retrieval of Authentic Documents for Reader-Specific Lexical Practice. In *Proceedings of the InSTIL/ICALL 2004 Symposium on Computer Assisted Learning*, Venice, Italy.

Collins-Thompson, K. and Callan, J. (2005). Predicting reading difficulty with statistical language models. *Journal of the American Society for Information Science and Technology*, 56(13):1448–1462.

DuBay, W. H. (2004). The Principles of Readability. Costa Mesa, California. Impact Information.

<center>— **Document Retrieval for Reading Practice** —</center>

Heilman, M., Zhao, L., Pino, J., and Eskenazi, M. (2008). Retrieval of Reading Materials for Vocabulary and Reading Practice. In *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications*, pages 80–88, Columbus, Ohio. Association for Computational Linguistics.

Miltsakaki, E. and Troutt, A. (2008). Real Time Web Text Classification and Analysis of Reading Difficulty. In *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications*, pages 89–97, Columbus, Ohio. Association for Computational Linguistics.

<center>55</center>

— Text Simplification —

Carroll, J., Minnen, G., Pearce, D., Canning, Y., Devlin, S., and Tait, J. (1999). Simplifying Text for Language-Impaired Readers. In *Proceedings of the Ninth Conference of the European Chapter of the Association for Computational Linguistics*, pages 269–270.

Inui, K., Fujita, A., Takahashi, T., Iida, R., and Iwakura, T. (2003). Text simplification for reading assistance: a project note. In *Proceedings of the second international workshop on Paraphrasing*, pages 9–16, Morristown, NJ, USA. Association for Computational Linguistics.

Lal, P. and Rüger, S. (2002). Extract-based Summarization with Simplification. In *Proceedings of the Workshop on Text Summarization at DUC 2002*.

Petersen, S. E. and Ostendorf, M. (2007). Text Simplification for Language Learners: A Corpus Analysis. In *Proceedings of Speech and Language Technology in Education (SLaTE2007)*, pages 69–72.

— Vocabulary Assistance —

Aist, G. (2001). Towards automatic glossarization: automatically constructing and administering vocabulary assistance factoids and multiple-choice assessment. *International Journal of Artificial Intelligence in Education*, 12:212 – 231.

Csomai, A. and Mihalcea, R. (2007). Linking Educational Materials to Encyclopedic Knowledge. In *Proceedings of the International Conference on Artificial Intelligence in Education (AIED 2007)*, Los Angeles, CA.

Mihalcea, R. and Csomai, A. (2007). Wikify!: linking documents to encyclopedic knowledge. In *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 233–242, New York, NY, USA. ACM.

Zesch, T., Gurevych, I., and Mühlhäuser, M. (2007). Analyzing and Accessing Wikipedia as a Lexical Semantic Resource. In Rehm, G., Witt, A., and Lemnitzer, L., editors, *Data Structures for Linguistic Resources and Applications*, pages 197–205. Gunter Narr, Tübingen.

Zesch, T., Müller, C., and Gurevych, I. (2008). Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary. In *Proceedings of LREC'08*.

— Spell Checking —

Heift, T. and Rimrott, A. (2008). Learner Responses to Corrective Feedback for Spelling Errors in CALL. *System*, 36:196–213.

Jurafsky, D. and Martin, J. H. (2008). *Speech and Language Processing*. Prentice Hall. 2nd edition.

Kukich, K. (1992). Techniques for automatically correcting words in text. *ACM Computing Surveys*, 24(4):377–439.

Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.

— Grammar Checking —

Atwell, E. S. (1987). How to detect grammatical errors in a text without parsing it. In *Proceedings of the third conference of the European chapter of the Association for Computational Linguistics*, pages 38–45, Morristown, NJ, USA. Association for Computational Linguistics.

Chodorow, M. and Leacock, C. (2000). An Unsupervised Method for Detecting Grammatical Errors. In *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics*, pages 140–147.

Chodorow, M., Tetreault, J., and Han, N.-R. (2007). Detection of Grammatical Errors Involving Prepositions. In *Proceedings of the 4th ACL-SIGSEM Workshop on Prepositions*, pages 25–30, Prague, Czech Republic. Association for Computational Linguistics.

Eeg-Olofsson, J. and Knutsson, O. (2003). Automatic grammar checking for second language learners - the use of prepositions. In *Proceedings of NoDaLiDa 2003*.

Lee, J. and Seneff, S. (2006). Automatic Grammar Correction for Second-Language Learners. In *Proceedings of INTERSPEECH 2006*, pages 1978–1981.

Lee, J. and Seneff, S. (2008). Correcting Misuse of Verb Forms. In *Proceedings of ACL-HLT-08*, pages 174–182.

Naber, D. (2003). A Rule-Based Style and Grammar Checker. Master's thesis, Technische Fakultät, Universität Bielefeld.

Nicholls, D. (1999). The Cambridge learner corpus - error coding and analysis. In *Summer Workshop on Learner Corpora*, Tokyo, Japan.

Sun, G., Liu, X., Cong, G., Zhou, M., Xiong, Z., Lee, J., and Lin, C.-Y. (2007). Detecting Erroneous Sentences using Automatically Mined Sequential Patterns. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*.

Wagner, J., Foster, J., and van Genabith, J. (2007). A Comparative Evaluation of Deep and Shallow Approaches to the Automatic Detection of Common Grammatical Errors. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 112–121.

— Dictionary Lookup —

Ferret, O. and Zock, M. (2006). Enhancing electronic dictionaries with an index based on associations. In *ACL '06: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL*, pages 281–288, Morristown, NJ, USA. Association for Computational Linguistics.

Tutoring Systems

Aleven, V., Koedinger, K. R., and Popescu, O. (2003). A Tutorial Dialog System to Support Self-Explanation: Evaluation and Open Questions. In *Proceedings of the 11th International Conference on Artificial Intelligence in Education (AIED 2003)*, pages 39–46.

Boyer, K., Phillips, R., Wallis, M., Vouk, M., and Lester, J. (2008). Learner Characteristics and Feedback in Tutorial Dialogue. In *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications*, pages 53–61, Columbus, Ohio. Association for Computational Linguistics.

Chi, M. T. H., Leeuw, N. D., Chiu, M.-H., and Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3):439–477.

Evens, M. and Michael, J. (2005). *One-on-One Tutoring by Humans and Computers*. Lawrence Erlbaum Associates.

Forbes-Riley, K. and Litman, D. J. (2005). Using bigrams to identify relationships between student certainness states and tutor responses in a spoken dialogue corpus. In *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue*, pages 87–96., Lisbon, Portugal.

Forbes-Riley, K. M., Litman, D., Huettner, A., and Ward, A. (2005). Dialogue-Learning Correlations in Spoken Dialogue Tutoring. In *Proceedings of the 12th International Conference on Artificial Intelligence in Education (AIED)*, pages 225–232, Amsterdam, The Netherlands.

Graesser, A., Wiemer-Hastings, K., Wiemer-Hastings, P., R, R. K., and the Tutoring Research Group U.o.M. (1999). AutoTutor: A simulation of a human tutor. *Cognitive Systems Research*, 1(1):35–51.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A., and Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers*, 36(2):180–192.

Graesser, A. C., Moreno, K. N., Marineau, J. C., Adcock, A. B., Olney, A. M., and Person, N. K. (2003). AutoTutor Improves Deep Learning of Computer Literacy: Is It the Dialog or the Talking Head? In *Proceedings of the 11th Conference on Artificial Intelligence in Education (AIED 2003)*, pages 47–54.

Graesser, A. C., VanLehn, K., Rosé, C. P., Jordan, P. W., and Harter, D. (2001). Intelligent Tutoring Systems with Conversational Dialogue. *AI Magazine*, 22(4):39–52.

57

Hausmann, R. G. and Chi, M. T. (2002). Can a Computer Interface Support Self-explaining? *Cognitive Technology Journal*, 7(1):4–15.

Lane, H. and VanLehn, K. (2005). Teaching the tacit knowledge of programming to novices with natural language tutoring. *Computer Science Education, Special issue on doctoral research in CS Education*, 15(3):183–201.

Litman, D. J., Rosé, C. P., Forbes-Riley, K., VanLehn, K., Bhembe, D., and Silliman, S. (2006). Spoken Versus Typed Human and Computer Dialogue Tutoring. *International Journal of Artificial Intelligence in Education*, 16(2):145–170.

Litman, D. J. and Silliman, S. (2004). ITSPOKE: An Intelligent Tutoring Spoken Dialogue System. In Susan Dumais, D. M. and Roukos, S., editors, *HLT-NAACL 2004: Demonstration Papers*, pages 5–8, Boston, Massachusetts, USA. Association for Computational Linguistics.

Person, N. K., Graesser, A. C., Bautista, L., Mathews, E., and the Tutoring Research Group (2001). Evaluating Student Learning Gains in Two Versions of AutoTutor. In *Proceedings of Artificial Intelligence in Education: AI-ED in the wired and wireless future*, pages 286–293.

Pon-Barry, H., Schultz, K., Bratt, E. O., Clark, B., and Peters, S. (2006). Responding to Student Uncertainty in Spoken Tutorial Dialogue Systems. *International Journal of Artificial Intelligence in Education*, 16(2):171–194.

VanLehn, K., Jordan, P. W., Rosé, C. P., Bhembe, D., Böttner, M., Gaydos, A., Makatchev, M., Pappuswamy, U., Ringenberg, M. A., Roque, A., Siler, S., and Srivastava, R. (2002). The Architecture of Why2-Atlas: A Coach for Qualitative Physics Essay Writing. In *ITS '02: Proceedings of the 6th International Conference on Intelligent Tutoring Systems*, pages 158–167, London, UK. Springer-Verlag.

Zinn, C., Moore, J. D., and Core, M. G. (2002). A 3-Tier Planning Architecture for Managing Tutorial Dialogue. In *ITS '02: Proceedings of the 6th International Conference on Intelligent Tutoring Systems*, pages 574–584, London, UK. Springer-Verlag.

Web 2.0 and Computer Supported Collaborative Learning

Bernhard, D. and Gurevych, I. (2008). Answering Learners' Questions by Retrieving Question Paraphrases from Social Q&A Sites. In *Proceedings of the 3rd Workshop on Innovative Use of NLP for Building Educational Applications, ACL 2008*, pages 44–52, Columbus, Ohio, USA.

— Quality of User-Generated Content —

Agichtein, E., Castillo, C., Donato, D., Gionis, A., and Mishne, G. (2008). Finding high-quality content in social media. In *WSDM '08: Proceedings of the international conference on Web search and web data mining*, pages 183–194, New York, NY, USA. ACM.

Druck, G., Miklau, G., and McCallum, A. (2008). Learning to Predict the Quality of Contributions to Wikipedia. In *Proceedings of the 'Wikipedia and Artificial Intelligence: An Evolving Synergy' Workshop at AAAI-08*.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438:900–901.

Jeon, J., Croft, W. B., Lee, J. H., and Park, S. (2006). A framework to predict the quality of answers with non-textual features. In *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 228–235, New York, NY, USA. ACM.

Kim, S.-M., Pantel, P., Chklovski, T., and Pennacchiotti, M. (2006). Automatically Assessing Review Helpfulness. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 423–430, Sydney, Australia. Association for Computational Linguistics.

Weimer, M. and Gurevych, I. (2007). Predicting the Perceived Quality of Web Forum Posts. In *Proceedings of the Conference on Recent Advances in Natural Language Processing (RANLP)*, pages 643–648.

Weimer, M., Gurevych, I., and Mühlhäuser, M. (2007). Automatically Assessing the Post Quality in Online Discussions on Software. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Companion Volume, Proceedings of the Demo and Poster Sessions*, pages 125–128, Prague, Czech Republic. Association for Computational Linguistics.

Electronic Career Guidance

Gurevych, I., Müller, C., and Zesch, T. (2007). What to be? - Electronic Career Guidance Based on Semantic Relatedness. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pages 1032–1039, Prague, Czech Republic. Association for Computational Linguistics.