Presentation at the Cross-Strait Conference on Educational and Psychological Testing,
October 19th-21st, 2018 at the Sun Moon Lake Teachers' Hostel, Nantou, Taiwan

Title: Automated scoring of Chinese compositions

Authors:
Magdalena Mo Ching Mok[1,2], Bor-Chen Kuo[1], Kai-Chih Pai[1], Jinxin Zhu[2], Sze Ming Lam[2], Kit Ying Tao[2], Tsz Wing Choi[2]

1 *National Taichung University of Education, Taiwan*

2 *The Education University of Hong Kong, Hong Kong*

Abstract:
Automated scoring of essays (AES) has earned a great deal of research attention in recent years. Proponents of AES (e.g., Deane, 2013; Gierl, Latifi, Lai, Boulais, & De Champlain, 2014; Ramineni & Williamson, 2013; Shermis, 2014) highlighted its efficiency in scoring, reporting, and feedback, as well as its capacity to augment human raters in the assessment of critical thinking and problem solving. Critics of AES (e.g., Condon, 2013; Les Perelman, 2014) on the other hand showed discontent on the assessment being confined to linguistic features at sentence- or word- levels rather than on the substantive issues of the writing contents. Despite mixed research findings on the effectiveness of AES, commercial packages such as e-rater® and IntelliMetric® are available for AES of essays written in the English language and AES has been applied to the scoring of commercial examinations as SAT, TOEFL. In contrast, nothing is yet in the market for AES in Chinese. This presentation aims to explore the possibility of using AES for Chinese essays, with the focus of providing timely formative feedback in support of self-directed learning of L1 students in Key Stage (KS) 2 (Primary 4 - 6) in Hong Kong in their learning of Chinese writing. The research team first built a corpus comprising 306 articles from 24 text books for KS2 students, totaling 433,134 Chinese characters. Next, 395 student essays were scored independently by teachers and two raters (the fourth grade undergraduate students majored in Chinese language). These scores were mapped onto a common Rasch scale using Facets software to obtain Rasch scores. Meanwhile, 62 indices were calculated by our AES system, developed on the basis of research at the National Taichung University of Education, using Latent Semantic Analysis (Liao, Chen, Kuo, & Pai, 2014). A prediction equation was developed using multiple regression on the Rasch scores by the AES indices; 12 indices were kept in the final equation with a R^2 of .470. Third, a new set of 2,795 essays from KS 2 students, totaling 816,981 Chinese characters was automatically rated by the AES system using the established equation. Based on the results, features were incorporated in the AES system to provide feedback to Chinese L1 learners on their essays.

References

Condon, W. (2013). Large-scale assessment, locally-developed measures, and automated scoring of essays: Fishing for red herrings? *Assessing Writing, 18* (1), 100-108.

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing, 18*, 7–24. doi: 10.1016/j.asw.2012.10.002

Gierl, M. J., Latifi, S., Lai, H., Boulais, A.-P., & De Champlain, A. (2014). Automated essay scoring and the future of educational assessment in medical education. *Medical Education, 48*(10), 950-962. doi:10.1111/medu.12517.

Les Perelman. (2014). When "the state of the art" is counting words. *Assessing Writing, 21*, 104-111.

Liao, C.-H., Chen, S.-M., Kuo, B.-C., Pai, K.-C. (2014). A Chinese vocabulary learning system: Latent semantic analysis approach. *International Journal of Innovative Computing, Information and Control, 10*(6), 2179-2191.

McNamara, D. S., & Graesser, A. C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P. M. McCarthy & C. Boonthum (Eds.), *Applied natural language processing: Identification, investigation, and resolution*. Hershey, PA: IGI Global.

Ramineni, C., & Williamson, D. (2013). Automated essay scoring: Psychometric guidelines and practices. *Assessing Writing, 18,* 25–39. doi: 10.1016/j.asw.2012.10.002

Shermis, M. D. (2014). State-of-the-art automated essay scoring: Competition, results, and future directions from a United States demonstration. *Assessing Writing, 20*, 53-76.

Acknowledgement

An early version of this presentation was made to the **A**dvanced **L**earning **T**heories, **T**echnologies, **A**pplications and **I**mpacts (ALTTAI) Consortium in Qinhuangdao, (Hebei, China) to take place on July 22nd through July 28, 2018