Automated Essay Assessment: An Evaluation on *PaperRater*'s Reliability from Practice

Nguyen Vi Thong
vithong1985@gmail.com
Graduate Institute of Linguistics
National Chung Cheng University, TAIWAN

ABSTRACT

From a perspective of a PaperRater user, the author attempts to investigate the reliability of the program. Twenty-four freshman students and one writing teacher at Dalat University - Vietnam were recruited to serve the study. The author also served as one scorer. The scores generated by PaperRater and the two human scorers were analyzed quantitatively and qualitatively. The statistical results indicate that there is an excellent correlation between the means of scores generated by three scorers. With the aid of SPSS and certain calculation, it is shown that PaterRater has an acceptable reliability which implies that the program can somehow assist in grading students' papers. The semi-structured interview at the qualitative stage with the teacher scorer helped point out several challenges that writing teachers might encounter when assessing students' prompts. From her perspective, it was admitted that with the assistance of PaperRater, the burden of assessing a bunch of prompts at a short time period would be much released. However, how the program can be employed by teachers should be carefully investigated. Therefore, this study provides writing teachers with pedagogical implications on how PaperRater should be used in writing classrooms. The study is expected to shed new light on the possibility of adopting an automated evaluation instrument as a scoring assistant in large writing classrooms.