Can Computer-Generated Linguistic Features Predict Second Language Students’ Writing Scores across Time?

Vahid Aryadoust  
Centre for English Language Communication, National University of Singapore  
*Corresponding Author’s E-mail: elcsva@nus.edu.sg

**Keywords:** Assessing writing; automated rating; automatic linear modeling; latent growth curve model; many-facet Rasch model

**EXTENDED ABSTRACT**

One of the most significant discussions in writing assessment literature is understanding the relationship between the quality of second language (L2) students’ texts as assessed by human raters and computer-generated linguistic features (Crossley, Weston, McLain Sullivan, & McNamara, 2011). Using human raters to judge the quality of L2 students’ essays is perceived to be a useful and relatively reliable rating method in writing assessments. Researchers have used a variety of holistic and analytic writing scales to assess students’ writing skill(s) and use the test scores for different purposes.

Despite its benefits, human rating has multiple limitations such as logistical constraints, cost inefficiency, and the need for double-marking to improve scores’ precision. To address these limitations, researchers have recently adapted automated rating engines. Computer-generated scoring or automated rating is time- and cost-efficient and recent developments in computer science have helped improve its precision significantly (Enright & Quinlan, 2010).

Although automatic and human rating have achieved relative success in writing assessment, there has been little discussion about the development of linguistic features and their links with writing quality over time. If evidence for developing linguistic features and their power to predict the growth of students’ writing quality is found, there will be much pedagogical and assessment use for them in academic writing. L2 writing development is not a well-defined science and it remains unexplored whether linguistic features could help researchers tease out and address the developmental features of texts.

The present study examines the link between the linguistic features as measured by Coh-Metrix and writing scores assigned by human raters over an academic semester.

**Methodology and Results**

**Participants**

This study uses the data from 116 first-year tertiary students enrolled in a paragraph writing module and aged between 18 and 20. They are enrolled in various disciplines including Business, Computer and Electrical Engineering, Geography, Real Estates, and Social Sciences. They performed six paragraph writing tasks at three time points (i.e., Pre-, Mid-, and Post-Course; two tasks per testing session) and their paragraphs were both marked by human raters and subjected to Coh-Metrix analysis.

**Generating Data**

The texts collected in the pre-, mid-, and post-course time points were marked by multiple raters on Content (i.e., the fulfillment of the task’s requirements), Organization (i.e., unity and cohesion in the paragraph), and Language (i.e., the richness and flexibility of vocabulary, grammar, and accuracy of mechanics). At each time point, 232 texts were collected (i.e., 116 students X 2 tasks), making 696 texts (i.e., 232 texts X 3 time points). Each text included one paragraph comprising between 180 to
230 words with the average of 210 words. Four contracted raters (CR)—who had taught the module for several semesters and used the analytical scale for rating before—marked 464 pre- and post-course texts and the four module tutors (MR) who taught the module marked 232 mid-course texts.

The pre- and post-course data were marked by the CR and the mid-course data were marked by the tutors teaching the module. I performed a number of many-facet Rasch model (MFRM) on the FACETS computer package to examine the quality and reliability of the generated data (Engelhard, 2012). In all, MFRM identified 28 misfits, which were re-marked by the researcher and coordinator.

**Development in Rater-Assigned Marks**
A Factor-of-Curves latent growth model (LGM) was used to examine students’ development in Content, Language, and Organization over time. The slope coefficients of Content, Language, and Organization indicated that these scores grow rather rapidly from point one to point two in all writing skills (Duncan, Duncan, & Strychker, 2006). The growth gradually deceased from point two to point three although it had a positive trend.

**Coh-Metrix Data**
The pre-, mid-, and post-course scripts were subjected to the Coh-Metrix analysis, Version 3, and specific information for each script at word, sentential, and discourse level was recorded.

**Automatic Linear Modelling (ALM)**
ALM is a data-mining approach to identify the relationships underlying the raw data. The goal of ALM is to use independent or predictive variables to predict the dependent variable. In this study, the dependent variable was set to be students’ writing scores—which were validated by the MFRM. The independent variables were the vocabulary sophistication, lexical diversity, syntactic sophistication, and cohesion statistics yielded in the Coh-Metrix analysis.

The forward stepwise ALM analysis on the pre-, mid-, and post-course data showed that the predictive power of computer-generated Coh-Metrix coefficients reduced over time. Ten (10) Coh-Metrix indices such as paragraph length, connectives, text readability, and hypernymy for verbs were able to explain 29.3% of the variance observed in the pre-Course data (Adjusted R$^2 = .293$).

However, none of these variables except text readability emerged in the Mid-Course data analysis. The ALM analysis on the Mid-Course data showed that 10 Coh-Metrix indices such as additive connectives and text readability explained 13.3% of variance (Adjusted R$^2 = .133$), leaving a big amount of the variance unexplained. Finally, the ALM analysis on the post-course data showed that 9.6% of the variance observed in the post-course data was explained by four Coh-Metrix indices such as hypernymy and syntactic simplicity (Adjusted R$^2 = .096$).

**Conclusion**
This study set out to investigate the predictive power of computer-generated Coh-Metrix indices over time. It was found that the power of these linguistic features dropped throughout the educational program, although they were still capable of predicting students’ paragraph writing proficiency by the end of the course. The findings are likely due to the nature of the Coh-Metrix indices. They have been developed on the basis of patterns in corpora in academic and informational texts. These corpora, however, do not reflect growth patterns of writers, specifically L2 writers. Therefore, although previous research has found the Coh-Metrix indices useful in predicting the observed variance in data, they may not be used for tracking L2 students’ growth.

To our knowledge, this is the first study which examines the predictive power of the Coh-Metrix indices, and therefore the findings should be further studied in future. The present study used lengthy paragraphs as units of investigation. Future research can use lengthier units of discourse such as essay to examine the predictive power of these indices over time.
References


