**Response-to-Text Prompts to Assess Students' Writing Ability: Using Natural Language Processing for Scoring Writing At-Scale**

Richard Correnti (PI), Diane Litman and Lindsay Clare Matsumura (Co-PIs)[1]

**Executive Summary:** Analytic writing in response to text is fundamental to academic success. In prior work we have argued for the importance of assessing students' analytic writing as a means for understanding progress toward outcomes espoused in the ELA Common Core State Standards (Correnti, Matsumura, Hamilton and Wang, 2013) and as a means for understanding measures of effective teaching (Correnti, Matsumura, Hamilton and Wang, 2012). We have argued, therefore, for the importance of our response-to-text assessment (RTA) because, 1) it is an essential skill for secondary and post-secondary success; 2) current assessments, where they assess writing at all, typically examine writing divorced from text (Hirsch and Pondiscio, 2010); 3) as assessment consortia build CCSS assessments, it is likely that "reading will be assessed through writing, making writing even more critical" in the curricula (Calkins, Ehrenworth & Lehman, 2012, p. 10); and 4) increased attention on analytic writing in schools will result in interventions to improve it, thus creating a need to assess progress in both writing instruction and in students' production of analytic writing. Recent advances in artificial intelligence offer a potential way forward through automated essay scoring (AES) of students' analytic writing at-scale. Furthermore, AES scores could also be used in formative assessments with the goal of improving students' writing drafts, but also, perhaps, the teaching of writing. However, to achieve this aim there is agreement that AES systems need to be able to score substantive features of writing and assess the validity of writing constructs (not just demonstrate inter-rater agreement (IRA) with human scores).

Our project built off of our prior work developing a response-to-text assessment ($RTA_{mvp}$) and generating human scores with strong IRA. We extended our work on the RTA by generating AES scores on two of the five dimensions (evidence and organization) comparable to human scores. Our AES system for scoring the *Evidence* and *Organization* dimensions of the RTA was built using techniques from two areas of the field of Artificial Intelligence: machine learning and natural language processing. Machine learning has highly effective methods for developing classification models from a given training sample by intelligent searches through a space of possible models. Natural language processing, in turn, has methods for transforming raw essay texts into a set of rubric-motivated features (i.e., independent variables) for model development. We have reported the ICCs we obtained between human and computer scoring of $RTA_{mvp}$ for each of the two dimensions — evidence (see Rahimi et al., 2014 for the specific features used to replicate human decision-tree) and organization (Rahimi et al., 2015). These ICCs were lower for the computer by human scores (.62 for evidence and .51 for organization) than they were between two human raters (.67 for evidence and .69 for organization). However, predictive validity analyses demonstrate no difference for AES and human scores in their sensitivity to writing instruction effects (manuscripts in process). In a recent IES proposal submission we asked for funding to extend our success in evidence and organization with the $RTA_{mvp}$ to a new form of the RTA – $RTA_{space}$ in order to understand how our initial work generalizes to a new form. We propose to examine whether researchers can use the two versions of the RTA as alternate forms (i.e., pre- and post-tests, interchangeably). We also proposed to develop a feedback system to students based on AES features used to score each dimension. Thus, we proposed examining whether our AES scoring can be transformed into an intervention for improving students' analytic writing.

[1] Graduate Student Researchers Elaine Wang and Zahra Rahimi were critical members of the research team. They were both highly involved in writing the IES proposal, with Elaine Wang listed as a Co-PI.