eRevise+RF: An Automated Writing Evaluation System to Support Text-based Argumentation and Revision

Zhexiong Liu[1], Tianwen Li[1], Elaine Lin Wang[2], Rip Correnti[3], Lindsay Clare Matsumura[3], Diane Litman[4]

[1,3,4]University of Pittsburgh, [2]RAND Corporation

[1]Presenter, [2]RAND Co-PI, [3]Co-PI, [4]PI

Automated Writing Evaluation (AWE) Systems

Writing is critical to learning and academic success (Graham & Perin, 2007). More recently, educational standards have begun to emphasize the significance of text-based argumentative writing in preparing students for college (e.g., Graham et al.,, 2015). This form of writing requires students to express higher-level thinking about texts, formulate arguments, and marshal solid evidence in support of their claims.

Research indicates that a significant number of young students find it challenging to construct compelling arguments. Offering formative feedback on preliminary versions of their assignments is crucial to enhance this aspect of writing (Graham et al., 2015).

However, students rarely receive substantive formative feedback on their writing for multiple reasons:

- Time Constraints: Providing formative feedback is time-consuming
- Uncertainty About Feedback: Being unsure about how to provide feedback to improve students' writing
- Surface-Level Focus: Focusing on surface-level features of students' writing

One approach to potentially easing the burden on teachers and increasing students' opportunities to receive substantive formative feedback is to leverage automated writing evaluation (AWE) systems. These systems combine automated essay scoring (AES) technologies with feedback on drafts of students' essays.

Current Project Goals

The goal of our current project is to develop and study an AWE system (**eRevise+RF**) for machine scoring the quality of students' revision aligned with formative feedback messages about their use of text evidence. Automated feedback messages for students in grades 5-7 will be generated to improve students' revision of text-based argument essays based on the automated scores.

Building upon our prior work, wherein we successfully developed an AWE system named **eRevise** to enhance students' incorporation of source text evidence into their argument writing, in current study we plan to:

- develop and establish the reliability and validity of new measures of revision quality in response to formative feedback on evidence use,
- use NLP to automate the scoring of revisions using these measures
- provide formative feedback messages to students based on the automated revision scoring and evaluate the utility of this feedback in improving student writing and revision.

Reference

Correnti, R., Matsumura, L. C., Wang, E., Litman, D., Rahimi, Z., & Kisa, Z. (2020). Automated scoring of students' use of text evidence in writing. *Reading Research Quarterly*, *55*(3), 493-520.

Graham, S., & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools: A report to Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education.

Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal*, *115*(4), 523–547.

Wang, E. L., Matsumura, L. C., Correnti, R., Litman, D., Zhang, H., Howe, E., ... & Quintana, R. (2020). eRevis (ing): Students' revision of text evidence use in an automated writing evaluation system. *Assessing Writing*, *44*, 100449.

Design of eRevise+RF

Response-to-text Assessment (RTA)

eRevise+RF is designed to score responses and provide feedback to students on the Response-to-Text Assessment (RTA). The RTA was developed to create a feasible means for assessing students' ability to reason about texts in their writing and use text evidence effectively to support their claims (See Figure 1).

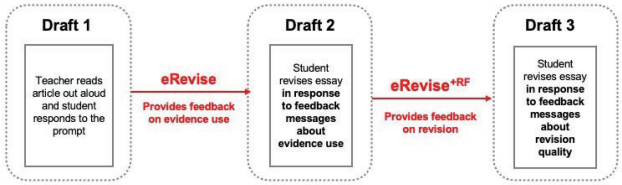

Figure 1. RTA & eRevise+RF

eRevise+RF System

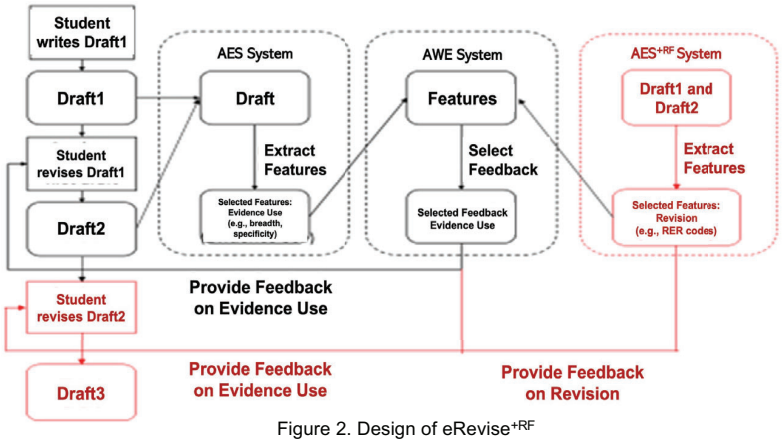

Figure 2. Design of eRevise+RF

AES System

- Provide automated essay scoring for draft 1 and draft 2.
- The scoring focuses on how students use evidence properly which evaluates both specificity and breadth of evidence use.

AWE System

- Based on two features extracted in AES system, e.g., breadth of evidence used.
- Provide automated feedback to students on drafts of their essays.
- Increase students' opportunities to revise their writing.

AES/AWE+RF System

- Automated assessment of students' revision quality (i.e., sentence-level revisions between draft 1 and draft 2).
- Assess holistic quality of revisions in terms of how well revision is aligned with feedback messages provided to the students on draft 1.
- Provide feedback to students on revision quality to help students revise for draft 3.

Results from Previous Research on eRevise

Beneficial to Students. The large majority of students (approximately 80%) indicated that they used 'a lot' or 'all' of the feedback they received to revise their essays and showed significant improvement in their evidence score from first to second drafts (Wang et al., 2020).

Beneficial for Teachers. Teachers appreciated the time saved from grading and the ability for students to receive timely feedback on their writing afforded by eRevise (Correnti et al., 2020), and also reported that the messages were aligned with their instructional goals.

Future Work

Generative AI to Provide Personalized Feedback

Individualized Writing Analysis: Generative AI will perform in-depth analyses of a student's writing, identifying specific evidence use and its improvement based on the individual's extracted features (Figure 2). The generative AI will offer personalized suggestions to address these concerns.

Tailored Recommendations and Suggestions: Generative AI will provide tailored feedback by understanding the student's skill level and the context of their writing. It will offer personalized suggestions to improve the evidence use based on the specific examples used in their drafts and help rank the most important improvement the student needs to concentrate on.

Fairness Consideration of Revision feedback

Bias Detection: Fairness in revision feedback involves detecting and addressing biases within the feedback provided by AI models. To ensure fairness, it's crucial to implement mechanisms that identify and mitigate biases in the feedback process. This will include monitoring for any biases related to race, gender, ethnicity, or cultural background in the suggestions or evaluations given to students.

Diversity and Inclusivity in Examples and References: Fair revision feedback will encompass a diverse range of examples and references. AI models used for feedback will suggest examples or references that represent a variety of perspectives and demographics, ensuring inclusivity and fairness, which can help students from different backgrounds feel equally represented and supported in their learning process.

Contact Information

Zhexiong Liu, CS Ph.D. Student
zhexiong@cs.pitt.edu

Tianwen Li, Education Ph.D. Student
tianwen.li@pitt.edu

Visit our webpage by scanning the QR code